Ph.D. Dissertation of Engineering

Improved Hessian-Free optimization for acoustic full waveform inversion

음향파 파형역산을 위한
개선된 헤시안-프리 최적화 기법

August 2017

Graduate School of Engineering
Seoul National University
Energy Systems Engineering Major

Shinwoong Kim

Abstract

Improved Hessian-Free optimization for acoustic full waveform inversion

Shinwoong Kim

Department of Energy Systems Engineering

The Graduate School

Seoul National University

Seismic full waveform inversion (FWI) is a method to reconstruct material properties of subsurface structures by minimizing the objective function based on residuals between modeled and observed seismic data. For seismic inverse problem, various kinds of optimization methods have been introduced. The truncated Newton method, also known as the Hessian-free (HF) optimization method, has been chosen to optimize large-scale inverse problems. The HF does not need to explicitly compute, store and invert the Hessian matrix. Instead of the Hessian matrix itself, the product of Hessian matrix and column vector is used for the linear conjugate-gradient loop during FWI process. To calculate the product of the Hessian matrix and column vector, the second-order adjoint (SOA) method or finite difference approximation (FDA) method has been widely used. The FDA is easy and intuitive to use in the linear conjugate-gradient method compared with SOA. The accuracy of FDA is dependent on not only the approximation interval but also the inversion settings, such as the model parameter, initial model, frequencies, etc.

To overcome dependency of HF optimization on the approximation method and inversion setting, an improved method is proposed for a stable HF optimization

method. The derivations of the improved method are based on not the FDA method but the limit of a function, which is independent of epsilon value. In other words, the improved HF method stably and accurately approximates the matrix-vector product of the Hessian matrix and column vector without any selection of epsilon value. In addition, computational cost of the improved HF optimization method is much lower than the conventional HF optimization method because additional construction and factorization of modeling operator are not needed during the linear conjugate-gradient method in the improved HF optimization method. To demonstrate the feasibility of the improve HF method, numerical examples for the Marmousi and acoustic Overthrust models are performed. Numerical examples indicate that the improved HF method shows better computational efficiency and stability than the conventional HF method without any degradation of inversion results.

Keyword: seismic full waveform inversion, truncated Newton method, Hessian-free method, acoustic wave equation

Student Number: 2012-24168

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1. Motivation

One of the most significant and popular areas in geophysics is inferring material properties of subsurface structures. Among several techniques used for the estimation of material properties, seismic full waveform inversion (FWI) has been extensively studied in oil and gas exploration, because it provides detailed information of subsurface parameters, such as P- and S-wave velocity, density, anisotropic properties and attenuation factor.

FWI reconstructs model parameters of subsurface structures by minimizing the objective function based on residuals between modeled and observed seismic data (Laily, 1983; Tarantola, 1984; Virieux and Operto, 2009). Among several optimization methods, the gradient-based method like the steepest-descent method has been widely applied for FWI because it is intuitive and easy to numerically implement. Laily (1983) and Tarantola (1984) proposed the adjoint-state method for calculating the gradient of objective function efficiently (Plessix, 2006; Shin and Min 2006). However, it has been widely known that the gradient-based optimization method is not suitable for solving nonlinear problems because of slow convergence rates and being easily stuck in local minima. To overcome these problems of the gradient-based method, several preconditioners have been introduced. Shin et al. (2001) and Choi et al. (2008) proposed using diagonal terms of the pseudo- or new pseudo-Hessian matrices as a preconditioner, respectively, instead of the Hessian matrix. These preconditioners compensate for the geometrical spreading effect of seismic wave propagations to some degree.

To solve the nonlinear problem, the second-order optimization method like the Newton-based method should be considered. However, for the large-scaled FWI, calculating, storing and inverting the Hessian matrix is too overburdensome with the present computer’s capacities. To overcome this problem, Brossier et al. (2010)

proposed the quasi-Newton method like *l*-BFGS for seismic FWI. The *l*-BFGS method does not require explicit calculation of the Hessian or its inverse but approximate the inverse matrix of Hessian by using some previous information of model parameters and gradients (Nocedal and Wright, 2006).

The truncated Newton method, also called the Hessian-Free (HF) optimization method, can be an attractive method for solving large-scaled and nonlinear FWI problem (Nash, 2000; Metivier et al., 2013; Wang et al., 2016). The explicit calculation of the Hessian matrix itself or its inverse is not required in FWI procedures based on the HF optimization. Only multiplication of the Hessian matrix with the column vector is needed for the linear conjugate-gradient procedure. The second-order adjoint-state method and the finite difference approximation (FDA) method have been proposed to calculate the product of the Hessian matrix and column vector (Nocedal and Wright, 2006; Metivier et al., 2013; Schiemenz et al., 2014). Between the second-order adjoint-state method and the FDA method, the FDA method has been popularly used and applied to calculate the product of the Hessian matrix and column vector because of its convenience and efficiency of implementation to the HF optimization algorithm (Wang et al., 2016). However, to accurately and stably approximate the matrix-vector product using the FDA method, it is essential to use an appropriate and small approximation interval for the FDA method. With too large an interval, the accuracy of the FDA method can be poor, whereas an extremely small interval can cause round-off errors. In both cases, FWI using the Hessian-free optimization method will fail to converge to the global minimum. The accuracy of the FDA can also be improved when the central FDA method is used instead of the forward FDA method. However, there exists a trade-off between computational cost and accuracy of the FDA because the central FDA method needs additional calculations to approximate the matrix-vector product of the Hessian matrix and column vector.

To accelerate the convergence rates of the HF method, an appropriate preconditioner can also be applied in the linear conjugate-gradient process. Pan et

al. (2016) applied the diagonal of the pseudo-Hessian matrix and the diagonal and pseudo-diagonal of the approximate Hessian matrix as preconditioners, and showed the faster convergence rates of the HF method with those preconditioners.

Although the HF optimization method does not require explicit calculation of the Hessian matrix, there still exists computational overburden for 3-D cases, because computational cost of the HF optimization method increases linearly depending on the number of sources. To accelerate the convergence rate of the HF method and to reduce computational cost, the simultaneous-source technique has been applied to FWI (Krebs et al., 2009; Ben-Hadj-Ali et al., 2011; Schuster et al., 2011; Jeong et al., 2013). Castellanos et al. (2015) proposed applying the simultaneous-source technique to the second-order optimization method, and showed that the simultaneous-source technique reduces computational cost and its convergence rate is faster than the individual-source technique.

1.2. Research Objectives

The HF optimization method approximates the matrix-vector product of the Hessian matrix and column vector by using the second-order adjoint-state method or the FDA method. Since the FDA method is more intuitive and efficient than the second-order adjoint-state method in the aspect of the numerical implementation, the FDA method became more popular among geophysicists. The only thing required for the FDA method is calculating differences between the steepest-descent directions with the perturbed model parameters.

However, to obtain accurate results using the FDA method, an appropriate interval needs to be chosen. With too large an interval, the approximation error occurs. In contrast, with too small an interval, the FDA method can be unstable because of round-off errors. Consequently, choosing an appropriate interval is one of the main issues to approximate the matrix-vector product of the Hessian matrix and column vector accurately and stably. In addition, although an appropriate interval is chosen in the beginning of FWI, it may not be appropriate for later iteration steps, because magnitudes of the steepest-descent directions vary and change as the FWI iteration proceeds.

To overcome these problems, a new approximation method using limit instead of the FDA method is proposed in this study. The new approximation method is not dependent on the approximation interval, and approximates the matrix-vector product of the Hessian matrix and column vector more accurately and stably than the FDA method without any degradation of FWI results. Furthermore, computational costs can be reduced, compared to the FDA method, which is because the new approximation method does not require constructing and factorizing the new modeling operator for the linear conjugate-gradient process.

1.3. Outline

Before describing the improved HF optimization method for seismic FWI, basic theories of the steepest-descent method and the conventional HF optimization method for seismic FWI will be discussed in Chapter 2. In addition to the optimization methods, several techniques like the preconditioning, weighting and source encoding, which can be also applied to the improved HF optimization, will also be introduced in Chapter 2. Introducing those basic theories, numerical tests obtained by the conventional HF optimization method are provided and analyzed to show several limitations of the conventional method. As a method to overcome the limitations of the FDA method, a new approximation method is introduced. In Chapter 3, the new approximation will be derived and its efficiency will be demonstrated. In Chapter 4, numerical examples for the Marmousi and acoustic Overthrust models will be presented to demonstrate improvement of stability, accuracy and computational cost of FWI using the improved HF optimization method.

Chapter 2. Review of forward and inverse theory

2.1. Forward modeling

The acoustic wave equation has been widely applied to simulate the wave propagations in seismic exploration. Although the real earth media have the elastic properties, the acoustic wave equation has been popularly used because of numerical simplicity and computational efficiency compared to the elastic wave equations. The 2D acoustic wave equation in the frequency domain can be written as follows:

$$-\frac{\omega^2}{\mathbf{c}^2}\mathbf{u} = \frac{\partial \mathbf{u}}{\partial x^2} + \frac{\partial \mathbf{u}}{\partial z^2} + \mathbf{f}\,, \tag{2.1}$$

where ω, $\mathbf{c}$, $\mathbf{u}$ and $\mathbf{f}$ indicate the angular frequency, P-wave velocity, pressure wavefields and source function in the frequency domain, respectively.

Several methods like the finite-difference method (Kelly et al., 1976), finite-element method (Zienkiewicz and Taylor, 2000) and pseudo-spectral method have been commonly used to approximate the acoustic wave equation. Among those methods, the finite-difference method is used to approximate the acoustic wave equation in this study because it is intuitive and numerically simple. The second-order spatial derivative terms in the right-hand side of equation (2.1) can be approximated by

$$\frac{\partial \mathbf{u}}{\partial x^2}+\frac{\partial \mathbf{u}}{\partial z^2} \approx \frac{\mathbf{u}(x+\Delta x,z)-2\mathbf{u}(x,z)+\mathbf{u}(x-\Delta x,z)}{\Delta x^2}+\frac{\mathbf{u}(x,z+\Delta z)-2\mathbf{u}(x,z)+\mathbf{u}(x,z-\Delta z)}{\Delta z^2}. \quad (2.2)$$

By substituting equation (2.2) into equation (2.1), the acoustic wave equation is expressed by the sparse linear system, which can be written as follows:

$$\mathbf{S}(\mathbf{x},\omega)\mathbf{u}(\mathbf{x},\omega)=\mathbf{f}(\mathbf{x},\omega), \quad (2.3)$$

where $\mathbf{S}$, so-called modeling operator, is complex-valued, symmetric and sparse matrix (Marfurt, 1984). By directly solving the linear problem (equation 2.3), the modeled pressure wavefield $\mathbf{u}(\mathbf{x},\omega)$ is obtained as follows:

$$\mathbf{u}(\mathbf{x},\omega)=\mathbf{S}^{-1}(\mathbf{x},\omega)\mathbf{f}(\mathbf{x},\omega). \quad (2.4)$$

Unlike seismic wave propagation in the real earth media, the modeled pressure wavefield contains undesired edge reflections generated due to finite size of the modeling domain. To suppress the undesired edge reflections, some boundary conditions have been proposed (Reynolds, 1978; Clayton and Engquist, 1977; Higdon, 1986; Shin, 1995; Cohen, 2002). In this study, the perfectly matched layer (PML) boundary condition (Cohen, 2002) is applied to suppress the artificially reflected waves.

2.2. Steepest-descent method for seismic FWI

The gradient-based method like the steepest-descent method is the one of the easiest and the most favorable methods to solve the optimization problem. The first-order optimization method like the steepest-descent method, which does not consider the Hessian matrix or its inverse, may not be appropriate for solving the nonlinear problem. It is well known that seismic FWI based on the steepest-descent method recovers only near-surface structures because the geometrical spreading effects are not compensated. To overcome this problem and to accelerate the convergence rate, the Gauss-Newton method has been used for seismic FWI, where the geometrical spreading effects are described by the approximate Hessian matrix. On the other hand, simplified preconditioners, such as the pseudo-Hessian and new pseudo-Hessian matrices, were introduced by Shin et al. (2001) and Choi et al. (2008). Those preconditioners have an effect of depicting the geometrical spreading effects to some degree.

The objective function for seismic FWI is based on the l_1-, l_2-, Huber or l_1/l_2 hybrid norms (Guitton and Symes, 2003; Pyun et al., 2009; Virieux and Operto, 2009). Among them, the l_2-norm has been widely used for seismic FWI, although it is not appropriate for noise-contaminated data like outlier. In this study, the objective function based on the l_2-norm is used.

Although the steepest-descent direction can be simply derived by using the adjoint-state method proposed by Laily (1983) and Tarantola (1984), computational cost almost increases linearly with the number of seismic sources. In case of using the source-encoding technique, computational cost of FWI can be almost independent of the number of seismic sources.

2.2.1 Inversion theory based on the steepest-descent method

In this study, the steepest-descent method is used to minimize the objective function based on the l_2-norm under the single source and frequency assumption, which is expressed by

$$E(\mathbf{m}) = \frac{1}{2}(\mathbf{Lu}-\mathbf{d})^t(\mathbf{Lu}-\mathbf{d})^* \ , \tag{2.5}$$

where the superscripts ^{t} and * indicate the transpose and conjugate operator, respectively. The column vectors $\mathbf{u}$, $\mathbf{d}$ and $\mathbf{m}$ indicate the modeled and observed pressure wavefield and model parameters for P-wave velocity, respectively. $\mathbf{L}$ is the matrix to project the whole dimension onto the receiver positions. To minimize the objective function with respect to the kth model parameter, taking partial derivative of the objective function with respect to the kth model parameter can be expressed as follows:

$$\nabla_k E = \frac{\partial E}{\partial m_k} = \mathrm{Re}\left[\left(\frac{\partial \mathbf{u}}{\partial m_k}\right)^t (\mathbf{Lu}-\mathbf{d})^*\right]. \tag{2.6}$$

Computing the partial derivative wavefields $\frac{\partial \mathbf{u}}{\partial m_k}$ occupies the most portions of computational cost during seismic FWI. To calculate equation (2.6) efficiently and fast, the adjoint method proposed by Laily (1983) and Tarantola (1984) has been commonly used in seismic FWI. Taking the partial derivative with respect to the kth model parameter, equation (2.3) is written as follows:

$$\frac{\partial \mathbf{S}}{\partial m_k}\mathbf{u} + \mathbf{S}\frac{\partial \mathbf{u}}{\partial m_k} = \mathbf{0}. \tag{2.7}$$

With equation (2.7), the partial derivative wavefields are derived efficiently as follows:

$$\frac{\partial \mathbf{u}}{\partial m_k} = \mathbf{S}^{-1}\left(-\frac{\partial \mathbf{S}}{\partial m_k}\mathbf{u}\right) = \mathbf{S}^{-1}\mathbf{v}_k. \tag{2.8}$$

Comparing equations (2.3) with (2.8), both equations require the matrix-vector product between the inversion of the modeling operator $\mathbf{S}^{-1}$ and the source vector $\mathbf{f}$ or virtual source vector $\mathbf{v}$, respectively. Substituting equation (2.8) into equation (2.6), the gradient of the k[th] model parameter can be expressed as follows:

$$\begin{aligned}\nabla_k E = \frac{\partial E}{\partial m_k} &= \mathrm{Re}\left[\left(\mathbf{S}^{-1}\mathbf{v}_k\right)^t (\mathbf{Lu}-\mathbf{d})^*\right] \\ &= \mathrm{Re}\left[\mathbf{v}_k^t\left(\mathbf{S}^{-1}\right)^t (\mathbf{Lu}-\mathbf{d})^*\right]\end{aligned}. \tag{2.9}$$

Consequently, the model parameter is updated by

$$m_k^{iter+1} = m_k^{iter} - \alpha_{iter}\nabla_k E, \tag{2.10}$$

where α_k is the step length, which can be chosen as a constant small value or estimated by the line search technique.

2.2.2 Preconditioner

Unlike the second-order optimization method, the Hessian matrix is not considered in the gradient-based method. As a result, the steepest-descent method focuses on inverting near-surface structures and its convergence rate is slow. To overcome these problems, the preconditioned steepest-descent method has been proposed in seismic FWI (Guitton et al., 2012). The preconditioner plays a role in compensating for the geometrical spreading effect to some degree, therefore it helps to accelerate the convergence rates of seismic FWI by inverting the deeper structures at the same time. To avoid heavy computational cost, a diagonal matrix like the pseudo-Hessian is preferred as a preconditioner in seismic FWI.

For the preconditioner, the pseudo-Hessian (Shin et al., 2001) and new pseudo-Hessian (Choi et al., 2008) matrices have been popularly used. In this study, the pseudo-Hessian is applied because it has proven to be effective in acoustic FWI. Followed by Shin et al. (2001), the k^{th} diagonal element of the pseudo-Hessian matrix is derived by using the virtual source as follows:

$$\left[\mathbf{H}_{diagonal\_pseudo}\right]_k = \mathbf{v}_k^t \mathbf{v}_k^* . \tag{2.11}$$

Accordingly, computing the pseudo-Hessian matrix does not require additional computational cost. The model parameter is updated as follows:

$$m_k^{iter+1} = m_k^{iter} - \alpha_{iter}\left[\mathbf{H}_{diagonal\_pseudo} + \lambda\mathbf{I}\right]_k^{-1} \nabla_k E , \tag{2.12}$$

where $\mathbf{I}$ is the identity matrix and λ is the damping factor to avoid the singularity and to control the depth scaling like the Levenberg-Marquardt method (Levenberg, 1944).

2.2.3 Simultaneous source method

These days, 3D seismic exploration has been essentially carried out in oil and gas exploration. However, 3D seismic FWI is not feasible yet due to heavy computational cost. In 3D case, the number of sources drastically increases compared to that of 2D case. As mentioned before, computational cost of seismic FWI increases almost linearly with the number of seismic sources. To overcome this problem, the source-encoding technique has been proposed and applied by many geophysicists to mitigate computational overburden in seismic FWI (Krebs et al., 2009; Ben-Hadj-Ali et al., 2011; Schuster et al., 2011; Jeong et al., 2013; Castellanos et al., 2015). The source-encoding technique makes it possible to compute the gradient summed over sources with the same computational cost as for a single source. In the source-encoding technique, super-source is defined as a linear combination of individual sources as follows:

$$\mathbf{f}_{super} = \sum_{ishot=1}^{nshot} \alpha_{ishot} \mathbf{f}_{ind}^{ishot}, \tag{2.13}$$

where $\mathbf{f}_{super}$, $\mathbf{f}_{ind}^{ishot}$, *nshot* and α_{ishot} indicate the super-source, individual source, the number of seismic sources and complex-valued source encoder, respectively. In this study, the complex-valued source encoder is randomly generated at every iteration and its property is described as follows:

$$Exp\left[\alpha_i^* \alpha_j\right] = \delta_{i,j}, \tag{2.14}$$

where Exp stands for the expectation over α (Castellanos et al., 2015). With the source encoder, the encoded, modeled and observed wavefield can be expressed

as follows:

$$\begin{aligned}\mathbf{u}_{super} &= \sum_{i=1}^{nshot} \alpha_i \mathbf{u}_i \\ \mathbf{d}_{super} &= \sum_{i=1}^{nshot} \alpha_i \mathbf{d}_i\end{aligned} . \tag{2.15}$$

With equation (2.15), the encoded objective function and encoded gradient of the model parameter are written as follows:

$$\begin{aligned}E_{super}(\mathbf{m}) &= \frac{1}{2}\left(\mathbf{L}\mathbf{u}_{super} - \mathbf{d}_{super}\right)^t \left(\mathbf{L}\mathbf{u}_{super} - \mathbf{d}_{super}\right)^* \\ \nabla_k E_{super}(\mathbf{m}) &= \mathrm{Re}\left[\left(\frac{\partial \mathbf{u}_{super}}{\partial m_k}\right)^t \left(\mathbf{L}\mathbf{u}_{super} - \mathbf{d}_{super}\right)^*\right]\end{aligned} . \tag{2.16}$$

Equation (2.16) shows that the gradient summed over sources can be efficiently computed by using the encoded modeled and observed wavefield with the cost of computing the gradient for a single source. Substituting equation (2.15) into equation (2.16) yields:

$$\nabla_k E_{super}(\mathbf{m}) = \mathrm{Re}\left[\begin{array}{l}\displaystyle\sum_{i=1}^{nshot} \alpha_i \alpha_i^* \left(\frac{\partial \mathbf{u}_i}{\partial m_k}\right)^t \left(\mathbf{L}\mathbf{u}_i - \mathbf{d}_i\right)^* + \\ \displaystyle\sum_{i=1}^{nshot}\sum_{j\neq i}^{nshot} \alpha_i \alpha_j^* \left(\frac{\partial \mathbf{u}_i}{\partial m_k}\right)^t \left(\mathbf{L}\mathbf{u}_j - \mathbf{d}_j\right)^*\end{array}\right], \tag{2.17}$$

where the second term in the right-hand side indicates the crosstalk term and the expectation over the crosstalk term becomes zero as FWI iteration proceeds.

As mentioned before, computational cost of seismic FWI depends on the number of forward modeling; the forward modeling depends on the number of seismic sources. In the individual source algorithm, the total number of forward

modeling needed is $2\times nshot$, because in each source, two times of the forward modeling are needed to calculate the modeled wavefield $\mathbf{u}$ and the back-propagated wavefield $\left(\mathbf{S}^{-1}\right)^{t}\left(\mathbf{Lu}-\mathbf{d}\right)^{*}$.

2.2.4 Weighting method

Lack of the low-frequency components in seismic data is the one of the main issues in seismic FWI. Because the low-frequency components of the seismic data contribute to construction of the long-wavelength structures in the early stage of FWI process. Without good initial models (i.e., sufficiently close to true model), FWI gets easily stuck in local minima if the low-frequency components of the seismic data are not available. Even though low-frequency components are available, low-frequency components of the gradient do not contribute to the model update because of smaller amplitudes compared to those of the high frequency components (Oh and Min, 2013). To overcome this problem, some weighting techniques were proposed by Ha et al. (2009), Jang et al. (2009) and Jeong et al. (2016) to emphasize amplitudes of low-frequency components.

In this study, the weighting technique proposed by Jeong et al. (2016) is applied to enhance contribution of low-frequency components for the model update. The weighting technique proposed by Jeong et al. (2016) modifies the objective function, and introduces the preconditioned gradient of the model parameter as follows:

$$E_{weight}\left(\mathbf{m}\right)=\sum_{i=1}^{nfreq}\sum_{j=1}^{nshot}\omega_1/\omega_i\max\left[\left|\mathbf{J}_{i,j}^{T}\left(\mathbf{L}\mathbf{u}_{i,j}-\mathbf{d}_{i,j}\right)^{*}\right|\right]E_{i,j}\left(\mathbf{m}\right)$$

$$P^{-1}\nabla E\left(\mathbf{m}\right)=\frac{\sum_{i=1}^{nfreq}\sum_{j=1}^{nshot}\omega_1/\omega_i\max\left[\left|\mathbf{J}_{i,j}^{T}\left(\mathbf{L}\mathbf{u}_{i,j}-\mathbf{d}_{i,j}\right)^{*}\right|\right]\mathbf{J}_{i,j}^{T}\left(\mathbf{L}\mathbf{u}_{i,j}-\mathbf{d}_{i,j}\right)^{*}}{\sum_{i=1}^{nfreq}\left[\sum_{j=1}^{nshot}\omega_1/\omega_i\max\left[\left|\mathbf{J}_{i,j}^{T}\left(\mathbf{L}\mathbf{u}_{i,j}-\mathbf{d}_{i,j}\right)^{*}\right|\right]diag\left(\mathbf{J}_{i,j}^{T}\mathbf{J}_{i,j}^{*}\right)+\lambda_i\right]} \quad , \quad (2.18)$$

where $nfreq$, ω_i, $\mathbf{J}$ and P indicates the number of frequencies, the i^{th} angular

frequency, the Jacobian matrix and the preconditioner, respectively. In equation (2.18), the weighting factor is inversely proportional to each frequency and emphasizes low-frequency components of the gradient only to construct long-wavelength structures. Figure 2.1 shows the workflow of FWI procedure using the steepest-descent method preconditioned by the pseudo-Hessian matrix.

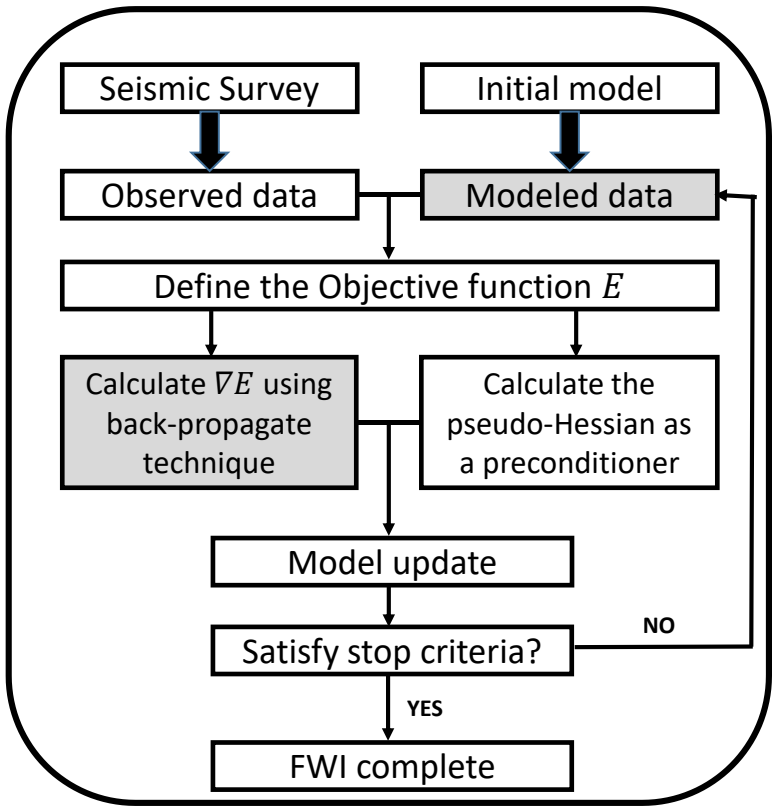

Figure 2.1 The workflow of the FWI algorithm using the preconditioned steepest-descent method. Light-grey box requires calculation of forward modeling once.

2.3. Hessian-Free optimization for seismic FWI

In recent years, the second-order optimization method, such as the quasi-Newton method like l-BFGS method and the truncated Newton method (i.e., the HF method), has attracted many applied mathematicians and engineers to optimize nonlinear problems. Unlike the gradient-based method, the second-order optimization method accounts for the Hessian matrix, which should be essentially considered to optimize the nonlinear problem. Many previous works have shown that the second-order optimization method gives much more reliable model parameters and much faster convergence rates than the gradient-based method (Brossier et al., 2010; Wang et al., 2016; Metivier et al., 2013; Schiemenz et al., 2014). Furthermore, for some FWI settings, the HF method outperforms the l-BFGS method (Metivier et al., 2013; Castellanos et al., 2015; Metivier et al., 2015). In this section, the basic theory of the HF optimization method for the seismic FWI will be discussed.

2.3.1 Review of the Hessian-Free optimization method

The second-order optimization method is derived from the second-order Taylor series approximation of the objective function with small model perturbation as follows:

$$E\left(\mathbf{m}_0+\Delta\mathbf{m}\right)\approx E\left(\mathbf{m}_0\right)+\Delta\mathbf{m}^t\nabla E\left(\mathbf{m}_0\right)+\frac{1}{2}\Delta\mathbf{m}^t\nabla^2 E\left(\mathbf{m}_0\right)\Delta\mathbf{m}, \quad (2.19)$$

where $\nabla^2 E\left(\mathbf{m}_0\right)$ is the Hessian matrix $\mathbf{H}$. From equation (2.19), the search direction $\Delta\mathbf{m}$ can be derived as follows:

$$\Delta\mathbf{m}=-\nabla^2 E\left(\mathbf{m}_0\right)^{-1}\nabla E\left(\mathbf{m}_0\right)=-\mathbf{H}\left(\mathbf{m}_0\right)^{-1}\nabla E\left(\mathbf{m}_0\right). \quad (2.20)$$

By using equation (2.20), the model parameters are updated during the iterative optimization procedure as follows:

$$\mathbf{m}^{iter+1}=\mathbf{m}^{iter}+\alpha^{iter}\Delta\mathbf{m}^{iter}. \quad (2.21)$$

However, explicit derivation, inversion and store of the Hessian matrix are too expensive for seismic FWI with the present computer's capacities, although the Hessian matrix is essential to optimize the seismic FWI. Because the HF optimization method does not require deriving, storing and inverting the Hessian matrix explicitly, the HF optimization method attracts many geophysicists' attentions. The linear conjugate-gradient method, which is employed inside the HF method, solves the linear problem described in equation (2.20) to derive the search directions of model parameter. In the linear conjugate-gradient process, the

algorithm never requires the exact Hessian matrix, but only requires the matrix-vector product of the Hessian matrix and column vector. By using this product, the search direction of model parameter is obtained without computational overburden.

Several methods have been proposed to efficiently calculate the matrix-vector product of the Hessian matrix and column vector. Metivier et al. (2013) proposed using the second-order adjoint method. Nocedal and Wright (2006) also proposed using the FDA. In this study, the FDA is employed because of its simplicity and efficient implementation (Wang et al., 2016). Figure 2.2 shows the workflow of seismic FWI algorithm using the HF optimization method. A dark-grey box in Figure 2.2 indicates that the forward modeling is performed twice for that procedure. Consequently, for a single shot and frequency, the number of forward modeling required to obtain the search direction of the model parameter using the HF optimization is $2+2\times CG$.

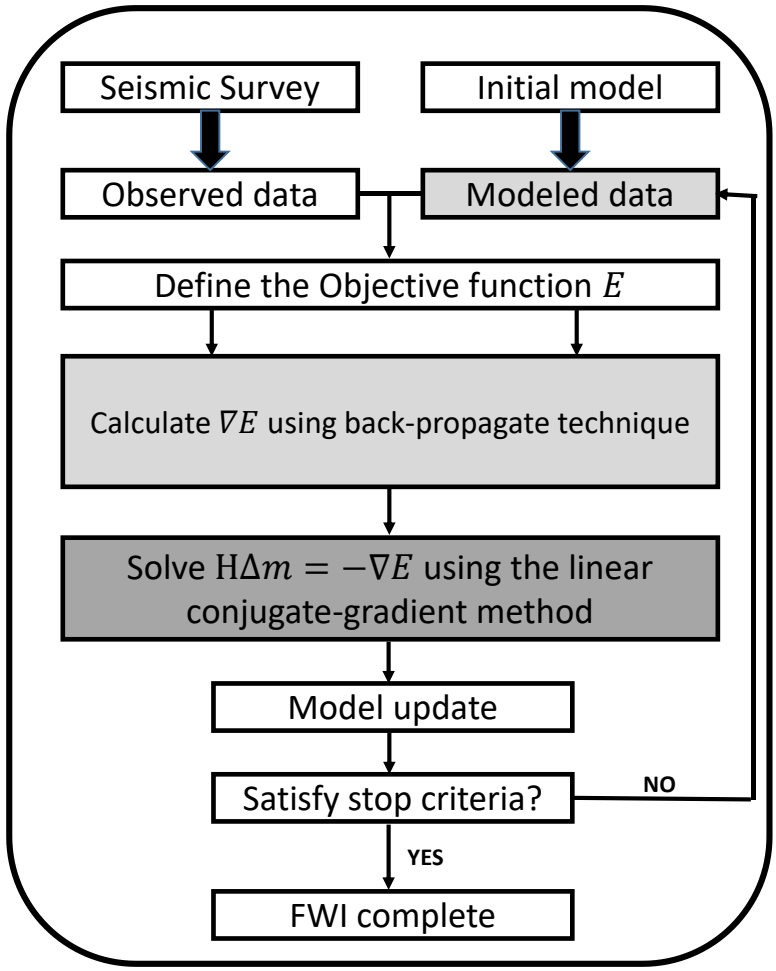

Figure 2.2 The workflow of FWI algorithm using the Hessian-free optimization method. Light-grey box requires calculation of forward modeling once. Dark-grey box requires calculation of forward modeling twice.

2.3.2 Linear conjugate-gradient method

As an iterative algorithm, the linear conjugate-gradient method has been widely employed to solve large-scale linear systems that are too heavy to solve directly. In the HF optimization method, the linear conjugate-gradient method is applied to solve equation (2.20) by minimizing

$$\mathbf{r}^k = \mathbf{b} - \mathbf{A}\mathbf{x}^k = -\nabla E - \mathbf{H}\Delta\mathbf{m}^k \quad . \tag{2.22}$$

The linear conjugate-gradient method begins with a good initial solution $\mathbf{x}^0$ of equation (2.20) if it is known. If not, the initial solution is set to be zero vector, and the other initial vectors are defined automatically as follows:

$$\begin{aligned} \mathbf{x}^0 &= \mathbf{0}; \\ \mathbf{r}^0 &= \mathbf{b}; \ . \\ \mathbf{p}^0 &= \mathbf{r}^0 \end{aligned} \tag{2.23}$$

After initialization, the linear conjugate-gradient method starts its own iteration as follows:

For icg = 0 ,⋯, ncgmax

$$
\begin{aligned}
\alpha^{icg} &= \frac{\left(\mathbf{r}^{icg}\right)^t \mathbf{r}^{icg}}{\left(\mathbf{p}^{icg}\right)^t \mathbf{A}\mathbf{p}^{icg}} \\
\mathbf{x}^{icg+1} &= \mathbf{x}^{icg} + \alpha^{icg}\mathbf{p}^{icg} \\
\mathbf{r}^{icg+1} &= \mathbf{r}^{icg} - \alpha^{icg}\mathbf{A}\mathbf{p}^{icg} \quad , \\
\beta^{icg} &= \frac{\left(\mathbf{r}^{icg+1}\right)^t \mathbf{r}^{icg+1}}{\left(\mathbf{r}^{icg}\right)^t \mathbf{r}^{icg}} \\
\mathbf{p}^{icg+1} &= \mathbf{r}^{icg+1} + \beta^{icg}\mathbf{p}^{icg}
\end{aligned}
\tag{2.24}
$$

End For CG iteration

where icg is the iteration number and ncgmax is the maximum iteration number. During the iterations, some stopping criteria for the loop are needed to improve computational efficiency of the truncated Newton method and to avoid using the non-positive definite (i.e., negative definite, positive semi-definite or negative semi-definite) matrix $\mathbf{A}$. The stopping criterion used in this study is expressed as follows:

$$\beta^{icg} > 1. \tag{2.25}$$

The other stopping criterion is the negative curvature test written by

$$\left(\mathbf{p}^{icg}\right)^t \mathbf{A}\mathbf{p}^{icg} > 0, \tag{2.26}$$

so that the linear conjugate-gradient method can be applied only when the matrix $\mathbf{A}$ is the positive definite matrix.

The matrix $\mathbf{A}$ in equation (2.24) indicates the Hessian matrix $\mathbf{H}$, and it is noticed that the matrix $\mathbf{A}$ itself (i.e., the Hessian matrix $\mathbf{H}$) is not needed inside

the linear conjugate-gradient loop. The matrix-vector product of $\mathbf{A}$ and $\mathbf{p}$ is needed to find a solution of equation (2.20). In other words, the Hessian matrix $\mathbf{H}$ itself, which is too expensive to derive, store and invert, is not needed although the second-order optimization method (i.e., the Hessian-free optimization) is applied to seismic FWI.

2.3.3 Matrix-vector product using the FDA method

The matrix-vector product of the Hessian matrix and column vector is the one of the most important processes in the HF optimization method. The FDA method is preferred to the second-order adjoint method because of its intuitive and convenient implementation. Nocedal and Wright (2006) proposed approximating the Hessian matrix with the FDA method based on the Taylor's series as follows:

$$\nabla E\left(\mathbf{m}_0+\varepsilon\mathbf{p}\right)=\nabla E\left(\mathbf{m}_0\right)+\varepsilon\nabla^2 E\left(\mathbf{m}_0\right)\mathbf{p}+O\left(\varepsilon^2\right), \tag{2.27}$$

$$\nabla^2 E\left(\mathbf{m}_0\right)\mathbf{p}=\mathbf{H}\left(\mathbf{m}_0\right)\mathbf{p}=\frac{\nabla E\left(\mathbf{m}_0+\varepsilon\mathbf{p}\right)-\nabla E\left(\mathbf{m}_0\right)}{\varepsilon}+O\left(\varepsilon\right). \tag{2.28}$$

Equation (2.28) shows that the matrix-vector product of $\mathbf{H}$ and $\mathbf{p}$ can be calculated using the forward FDA method with an appropriate interval ε. The approximation cost is just for computing additional gradient $\nabla E\left(\mathbf{m}_0+\varepsilon\mathbf{p}\right)$ at every linear conjugate-gradient iteration. Note that computing the additional gradient $\nabla E\left(\mathbf{m}_0+\varepsilon\mathbf{p}\right)$ requires the forward modeling twice with the new modeling operator $\mathbf{S}\left(\mathbf{m}_0+\varepsilon\mathbf{p}\right)$ and the approximation error is $O\left(\varepsilon\right)$. The matrix-vector product can also be approximated by the backward or central FDA method as follows:

$$\mathbf{H}\left(\mathbf{m}_0\right)\mathbf{p}=\frac{\nabla E\left(\mathbf{m}_0\right)-\nabla E\left(\mathbf{m}_0-\varepsilon\mathbf{p}\right)}{\varepsilon}+O\left(\varepsilon\right), \tag{2.29}$$

$$\mathbf{H}\left(\mathbf{m}_0\right)\mathbf{p}=\frac{\nabla E\left(\mathbf{m}_0+\varepsilon\mathbf{p}\right)-\nabla E\left(\mathbf{m}_0-\varepsilon\mathbf{p}\right)}{2\varepsilon}+O\left(\varepsilon^2\right). \tag{2.30}$$

The backward FDA method requires the same cost as the forward FDA method does, because the backward FDA method requires to derive the additional gradient $\nabla E\left(\mathbf{m}_0 - \varepsilon\mathbf{p}\right)$. Unlike the backward and forward FDA methods, the central FDA method requires computing two additional gradients (i.e., $\nabla E\left(\mathbf{m}_0 + \varepsilon\mathbf{p}\right)$ and $\nabla E\left(\mathbf{m}_0 - \varepsilon\mathbf{p}\right)$) at every iteration of the linear conjugate-gradient method. However, in the aspect of accuracy, it is obvious that the approximation error of the central FDA method is much smaller than those of the forward and backward FDA methods as shown in equations (2.28), (2.29) and (2.30). Therefore, there exists trade-off between accuracy and computational efficiency. Table 2.1 summarizes computational cost and accuracy for the FDA methods. Computational cost is computed by the number of the forward modeling required to calculate the model parameter update, and each gradient requires the forward modeling twice for calculating the modeled wavefields and back-propagated wavefields. The procedure of the linear conjugate-gradient algorithm considering the FDA method and stopping criteria is expressed as follows:

For icg = 0 ,$\cdots$, ncgmax

Calculate $\mathbf{Ap}^{icg}$ using FDA method

Check $\left(\mathbf{p}^{icg}\right)^{t}\mathbf{Ap}^{icg} > 0$

$$\alpha^{icg} = \frac{\left(\mathbf{r}^{icg}\right)^{t}\mathbf{r}^{icg}}{\left(\mathbf{p}^{icg}\right)^{t}\mathbf{Ap}^{icg}}$$

$$\mathbf{x}^{icg+1} = \mathbf{x}^{icg} + \alpha^{icg}\mathbf{p}^{icg}$$

$$\mathbf{r}^{icg+1} = \mathbf{r}^{icg} - \alpha^{icg}\mathbf{Ap}^{icg} \quad . \quad (2.31)$$

$$\beta^{icg} = \frac{\left(\mathbf{r}^{icg+1}\right)^{t}\mathbf{r}^{icg+1}}{\left(\mathbf{r}^{icg}\right)^{t}\mathbf{r}^{icg}}$$

check β

$$\mathbf{p}^{icg+1} = \mathbf{r}^{icg+1} + \beta^{icg}\mathbf{p}^{icg}$$

End For CG iteration

Table 2.1 Computational amount and accuracy of the forward, backward and central FDAs for $\mathbf{Hp}$ approximation. Computational amount is computed based on the number of forward modeling and conjugate gradient loops needed to compute the model parameter update.

	Approximation	Cost	Accuracy
Forward FDA	$\frac{\nabla E(\mathbf{m}_0+\varepsilon\mathbf{p})-\nabla E(\mathbf{m}_0)}{\varepsilon}$	$2+2CG$	$O(\varepsilon)$
Backward FDA	$\frac{\nabla E(\mathbf{m}_0)-\nabla E(\mathbf{m}_0-\varepsilon\mathbf{p})}{\varepsilon}$	$2+2CG$	$O(\varepsilon)$
Central FDA	$\frac{\nabla E(\mathbf{m}_0+\varepsilon\mathbf{p})-\nabla E(\mathbf{m}_0-\varepsilon\mathbf{p})}{2\varepsilon}$	$2+4CG$	$O(\varepsilon^2)$

2.3.4 Preconditioned Hessian-Free optimization

In section 2.2.2, the preconditioned steepest-descent method is discussed. In this section, in order to enhance convergence rates of the linear conjugate-gradient method in the HF optimization, the preconditioned HF optimization algorithm is introduced. The cost of the Hessian-free optimization method depends on the number of iterations for application of the linear conjugate-gradient method. Some preconditioners for the HF optimization method were proposed and investigated by Pan et al. (2016). The diagonal of the pseudo-Hessian matrix, diagonal of the Gauss-Newton Hessian matrix and pseudo-diagonal of the Gauss-Newton Hessian matrix were introduced as a preconditioner to accelerate the convergence rate of the HF optimization. In this study, as discussed in the previous section 2.2.2, the diagonal of the pseudo-Hessian matrix as a preconditioner is applied to accelerate the Hessian-free optimization method. The procedure of the HF optimization using the diagonal of the pseudo-Hessian matrix can be expressed as follows:

Initialization: $\mathbf{x}^0 = \mathbf{0},\ \mathbf{r}^0 = \mathbf{b}, \mathbf{p}^0 = \mathbf{P}^{-1}\mathbf{r}^0$

For icg = 0 ,$\cdots$, ncgmax

Calculate $\mathbf{Ap}^{icg}$ using FDA method

Check $\left(\mathbf{p}^{icg}\right)^t \mathbf{Ap}^{icg} > 0$

$$\alpha^{icg} = \frac{\left(\mathbf{r}^{icg}\right)^t \left(\mathbf{P}^{-1}\mathbf{r}^{icg}\right)}{\left(\mathbf{p}^{icg}\right)^t \mathbf{Ap}^{icg}}$$

$$\mathbf{x}^{icg+1} = \mathbf{x}^{icg} + \alpha^{icg}\mathbf{p}^{icg}$$

$$\mathbf{r}^{icg+1} = \mathbf{r}^{icg} - \alpha^{icg}\mathbf{Ap}^{icg}$$

$$\beta^{icg} = \frac{\left(\mathbf{r}^{icg+1}\right)^t \left(\mathbf{P}^{-1}\mathbf{r}^{icg+1}\right)}{\left(\mathbf{r}^{icg}\right)^t \left(\mathbf{P}^{-1}\mathbf{r}^{icg}\right)}$$

check β

$$\mathbf{p}^{icg+1} = \mathbf{P}^{-1}\mathbf{r}^{icg+1} + \beta^{icg}\mathbf{p}^{icg}$$

End For CG iteration (2.32)

Figure 2.3 shows the workflow of the preconditioned HF optimization. As mentioned in Figure 2.2, a dark-grey box in Figure 2.3 indicates that the forward modeling is performed twice for that procedure.

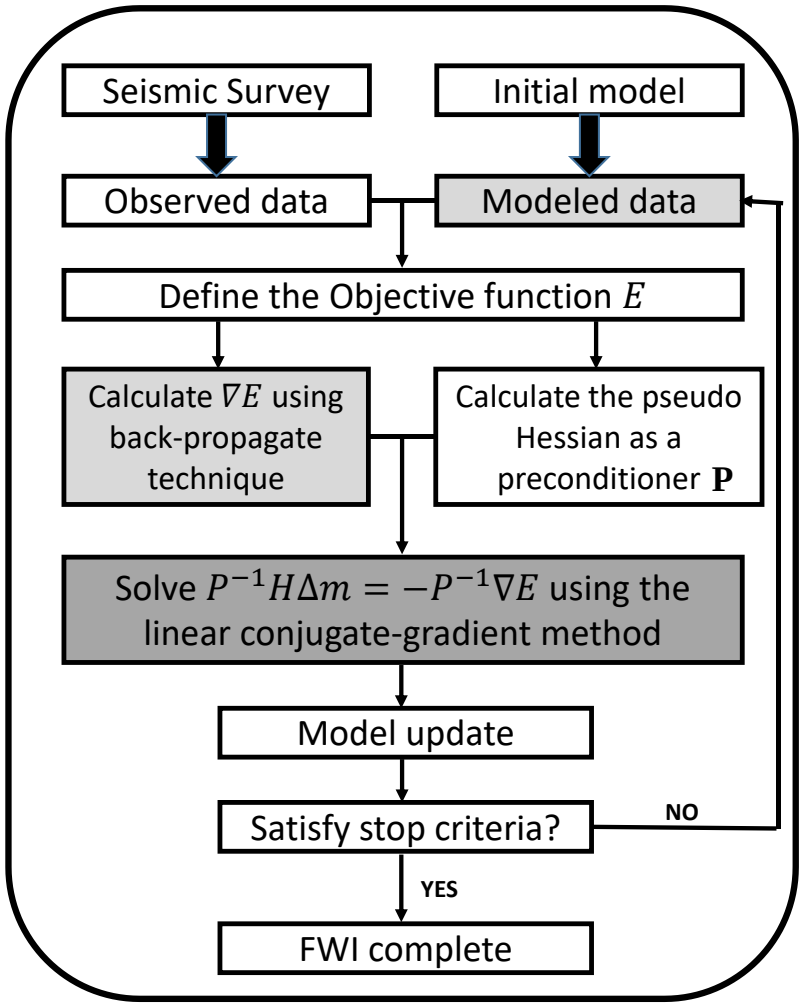

Figure 2.3 The workflow of the FWI algorithm using the preconditioned Hessian-free optimization method. Light-grey box requires calculation of forward modeling once. Dark-grey box requires calculation of forward modeling twice.

Chapter 3. Improved Hessian-Free optimization

3.1. Analysis of the Hessian approximation

The HF optimization method solves the nonlinear problem like the seismic FWI without full information of the Hessian matrix itself or its inverse. Instead, it requires the matrix-vector product of the Hessian matrix and column vector, which is approximated by the FDA method. Therefore, the accuracy and stability of the Hessian matrix information depend on those of the matrix-vector product approximation. To approximate the matrix-vector product, the forward FDA method has been preferred to the central FDA method because of computational efficiency. With an appropriate interval ε obtained by the trial and error method, the FDA method approximates the matrix-vector product reasonably. To investigate the stability and accuracy of the Hessian approximation according to the FWI settings, numerical examples for synthetic data will be presented in this section.

3.1.1 Analysis of the forward and central FDA methods

As described in Chapter 2, the product of the Hessian matrix and column vector is approximated by using the FDA method (equations 2.29 and 2.30). To analyze the accuracy and stability of the Hessian approximation based on the forward and central FDA methods, numerical tests for the Marmousi model are performed. Figure 3.1 shows the true and initial models for P-wave velocity and the FWI settings are described in Table 3.1. To avoid the multi-parameter problems, density is assumed to be homogeneous. Figures 3.2 and 3.3 show the matrix-vector products of the Hessian matrix and column vector, $\mathbf{Hp}$, computed by using forward and central FDA methods for various intervals ε ranging from 0.1 to 0.00001. As shown in Figure 3.3, the $\mathbf{Hp}$ approximation using the central FDA method gives consistent results for intervals ranging from 0.1 to 0.0001. However, the $\mathbf{Hp}$ approximation using the forward FDA method is not stable for an interval ε of 0.0001. In this case, the HF optimization method will fail to solve seismic FWI because of the wrong $\mathbf{Hp}$ approximation.

As mentioned in equations (2.29) and (2.30), it has been widely known that the central FDA method gives more accurate and stable approximation compared to the forward FDA method. In the aspect of the stability and accuracy, the central FDA method should be implemented to approximate the matrix-vector product of the Hessian matrix and column vector rather than the forward FDA method. However, most of the Hessian-free optimization studies have been developed based on the forward FDA method to decrease the number of forward modeling, which is directly related to computational cost, as shown in Table 2.1.

Table 3.1 Inversion settings used for numerical tests of the Marmousi model.

Model size	No. of shot	No. of receiver	Shot interval	Recording time	Maximum Frequency	Minimum Frequency
9.2 km **× 3.26 km**	150	450	0.06 km	8 sec	10 Hz	0.25 Hz

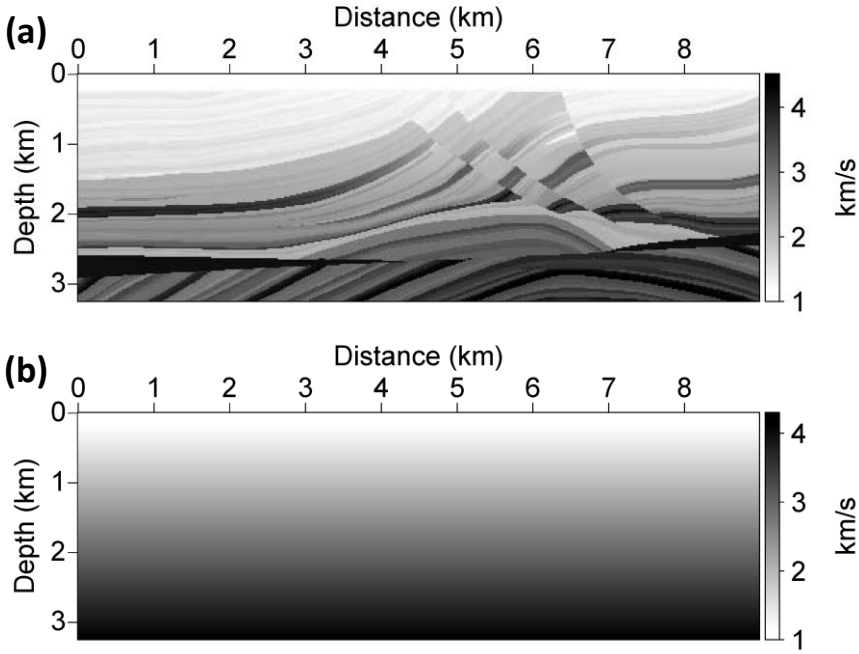

Figure 3.1 P-wave velocity models for numerical tests to obtain the approximation of $\mathbf{Hp}$ by using the forward and central FDA methods: (a) the true Marmousi and (b) linearly-increasing initial models.

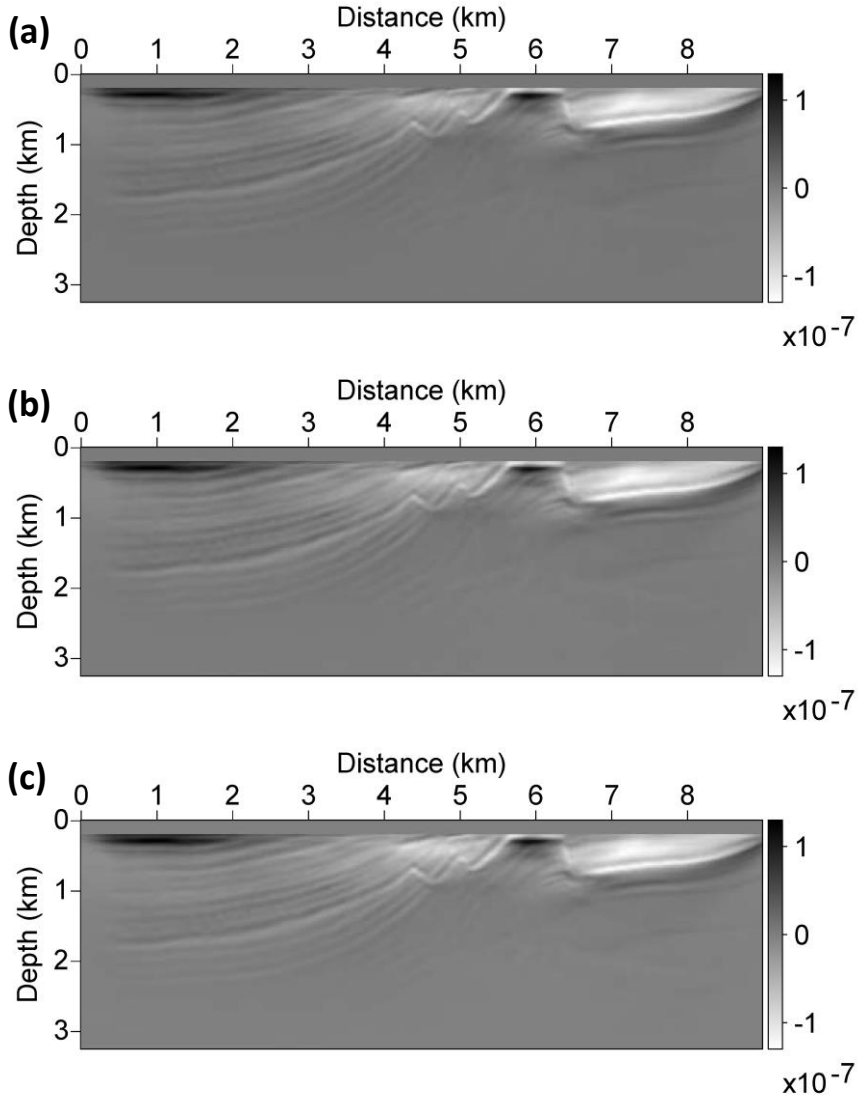

Figure 3.2 The matrix-vector product of the Hessian matrix and column vector **Hp** obtained using the forward FDA method with intervals of (a) 0.1, (b) 0.01, (c) 0.001, (d) 0.0001 and (e) 0.00001 for the Marmousi model.

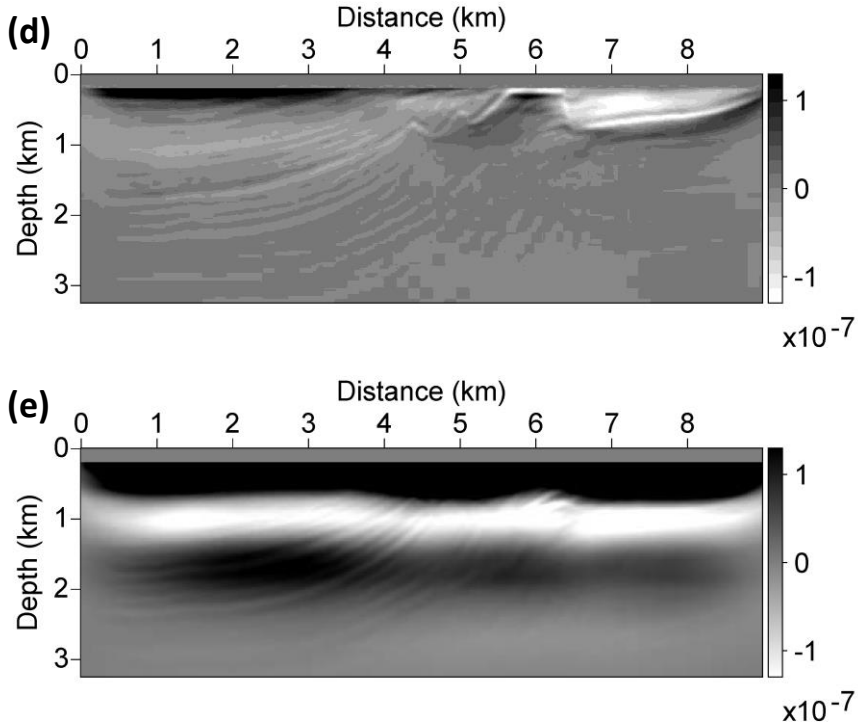

Figure 3.2 (Continued)

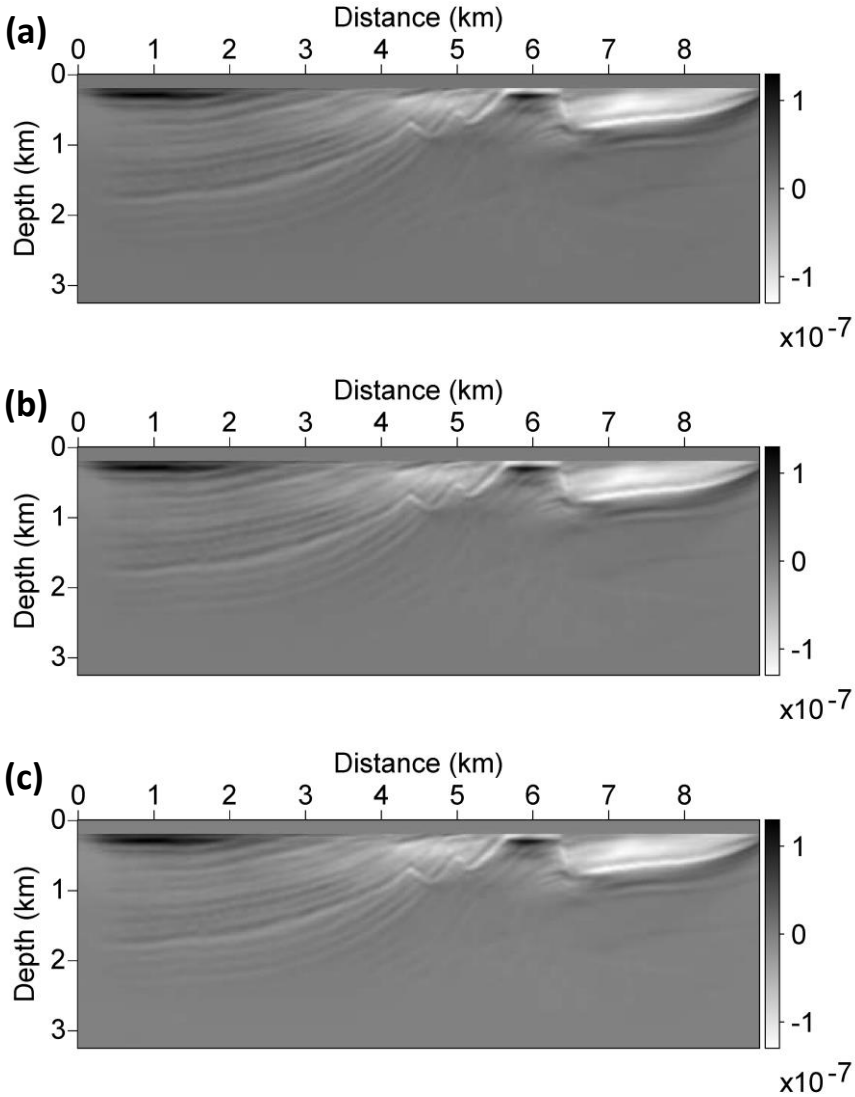

Figure 3.3 The matrix-vector product of the Hessian matrix and column vector **Hp** obtained using the central FDA method with intervals of (a) 0.1, (b) 0.01, (c) 0.001, (d) 0.0001 and (e) 0.00001 for the Marmousi model.

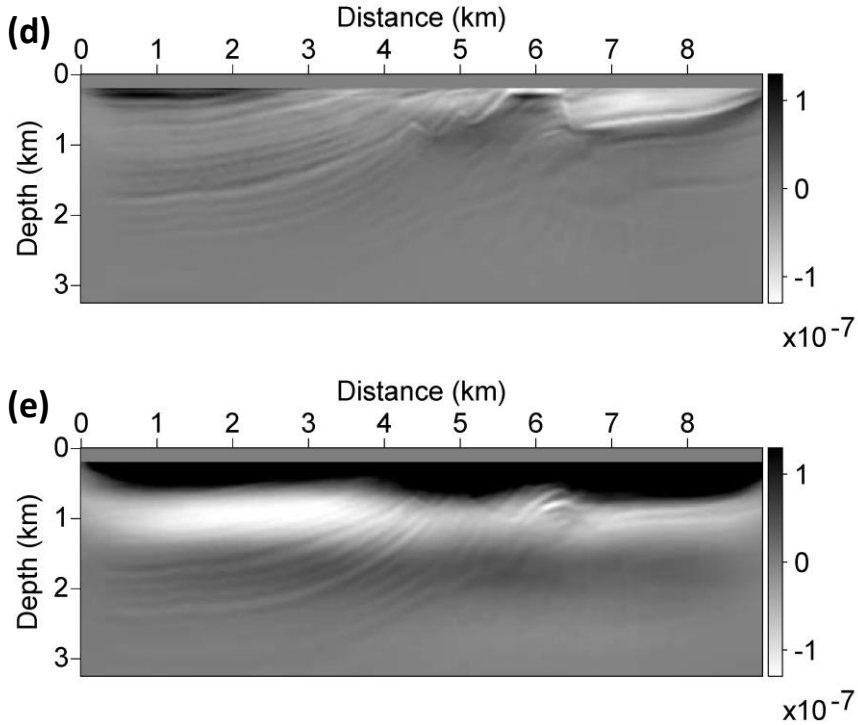

Figure 3.3 (Continued)

3.1.2 Analysis of frequency dependency

Followed by the general inverse theory, equation (2.20) can be rewritten as follows:

$$\mathbf{H}(\mathbf{m}_0)\Delta\mathbf{m} = -\nabla E(\mathbf{m}_0). \tag{3.1}$$

Note that the Hessian matrix and the steepest-descent direction of model parameter are composed of the summation over frequencies (equation 3.2) in the frequency domain as follows:

$$\sum_{i=1}^{nfreq} \mathbf{H}_i \Delta\mathbf{m} = -\sum_{i=1}^{nfreq} \nabla E_i \ . \tag{3.2}$$

In the same manner, the Hessian approximation in equation (2.28) is composed of the summation over frequencies, which can be expressed as follows:

$$\begin{aligned}
&\sum_{i=1}^{nfreq} \mathbf{H}_i\mathbf{p} = \mathbf{H}_1\mathbf{p} + \mathbf{H}_2\mathbf{p} + \cdots + \mathbf{H}_{nfreq}\mathbf{p} \\
&\approx \frac{\nabla E_1(\mathbf{m}_0 + \varepsilon\mathbf{p}) - \nabla E_1(\mathbf{m}_0)}{\varepsilon} + \frac{\nabla E_2(\mathbf{m}_0 + \varepsilon\mathbf{p}) - \nabla E_2(\mathbf{m}_0)}{\varepsilon} + \cdots . \\
&= \sum_{i=1}^{nfreq} \frac{\nabla E_i(\mathbf{m}_0 + \varepsilon\mathbf{p}) - \nabla E_i(\mathbf{m}_0)}{\varepsilon}
\end{aligned} \tag{3.3}$$

With the model perturbation vector $\mathbf{p}$, the first-order partial derivative of the objective function $\nabla E_i(\mathbf{m}_0 + \mathbf{p})$, (i.e., the steepest-descent direction of model parameter), have different sensitivity depending on frequency. Consequently, the interval ε can be different depending on frequency. Figures 3.4, 3.5 and 3.6 show

approximations of $\mathbf{H}_i\mathbf{p}$ using the forward FDM with various intervals for frequencies of 1.5, 2.5, and 5.5 Hz. The approximations of $\mathbf{H}_i\mathbf{p}$ in Figures 3.5 and 3.6 show similar results regardless of the interval ε, while the approximation of $\mathbf{H}_i\mathbf{p}$ obtained using ε of 0.0001 in Figure 3.4d shows different patterns compared to those for ε of 0.1, 0.01 and 0.001. To investigate the approximations more precisely, depth profiles recorded at 6.9 km are shown in Figures 3.7, 3.8 and 3.9. Comparing the depth profiles, the differences between the approximations decrease as the frequencies increase. It is obvious that the accuracy and stability of the $\mathbf{H}_i\mathbf{p}$ approximation depend on the interval ε. In other words, an interval, which is appropriate for a certain frequency, may fail in approximating $\mathbf{H}_i\mathbf{p}$ in other frequencies.

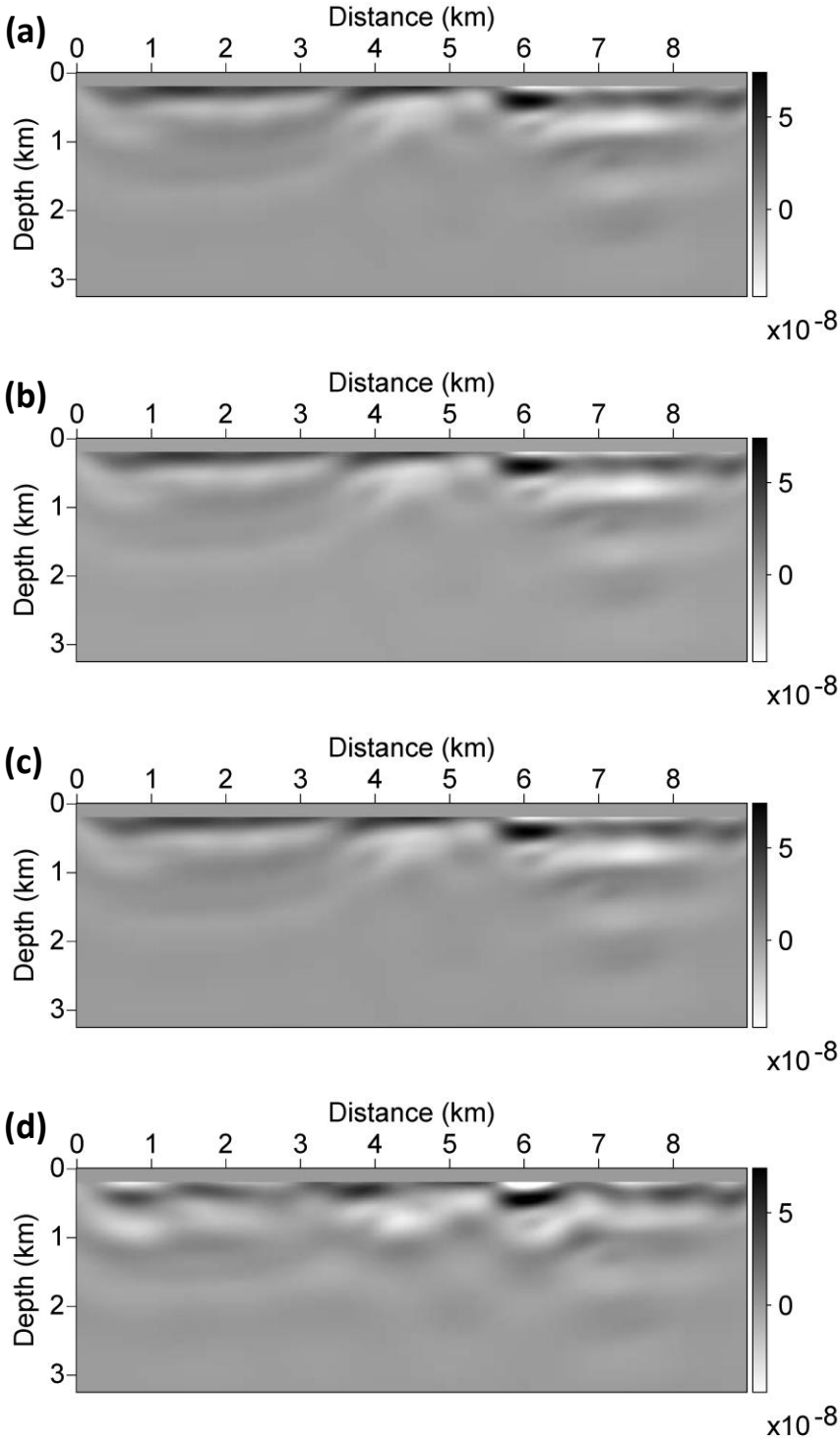

Figure 3.4 The matrix-vector product of the Hessian matrix and column vector $\mathbf{H}_i\mathbf{p}$ at 1.5 Hz obtained using the forward FDA method with intervals of (a) 0.1, (b) 0.01, (c) 0.001 and (d) 0.0001 for the Marmousi model.

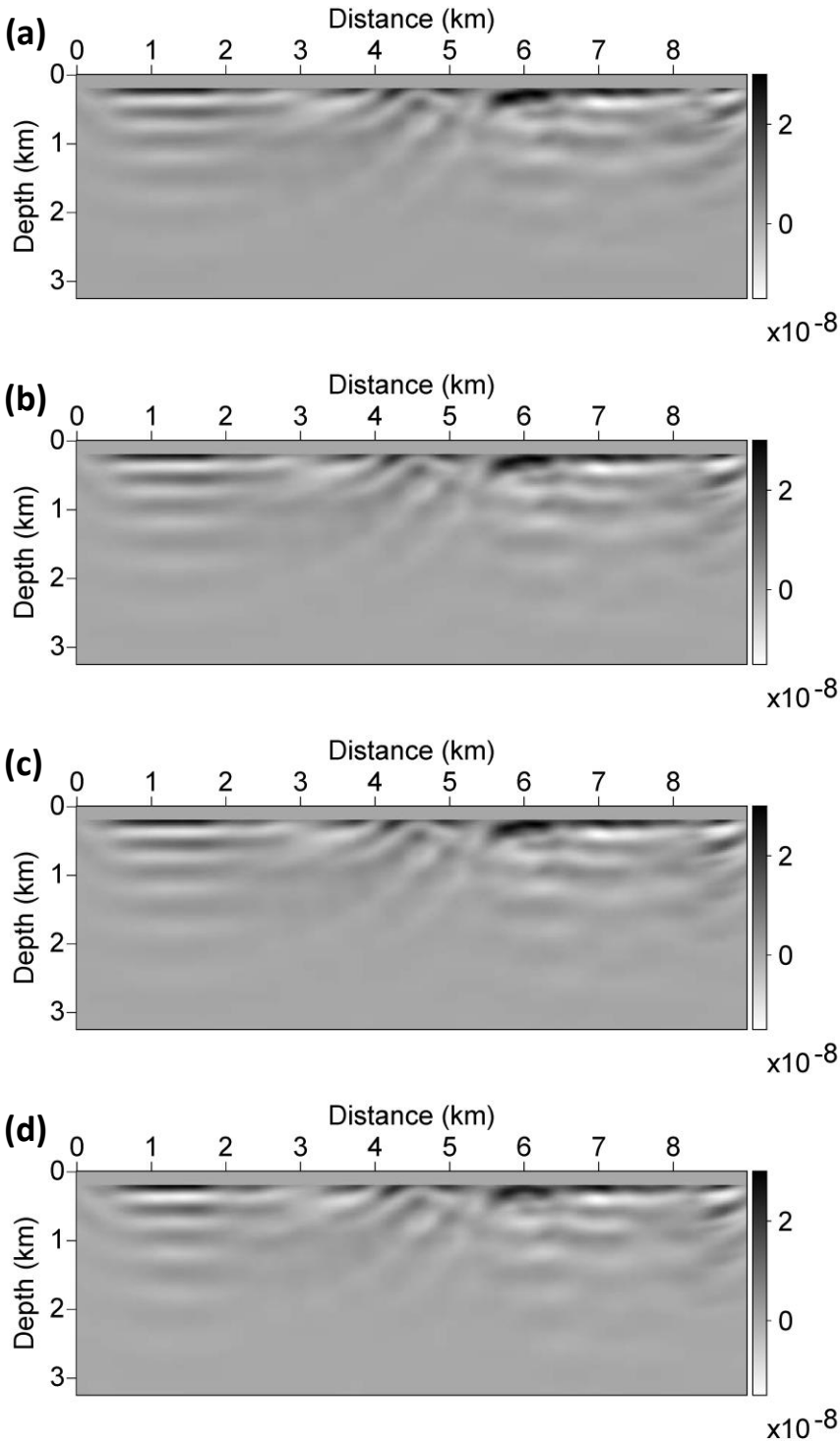

Figure 3.5 The matrix-vector product of the Hessian matrix and column vector $\mathbf{H}_i\mathbf{p}$ at 2.5 Hz obtained using the forward FDA method with intervals of (a) 0.1, (b) 0.01, (c) 0.001 and (d) 0.0001 for the Marmousi model.

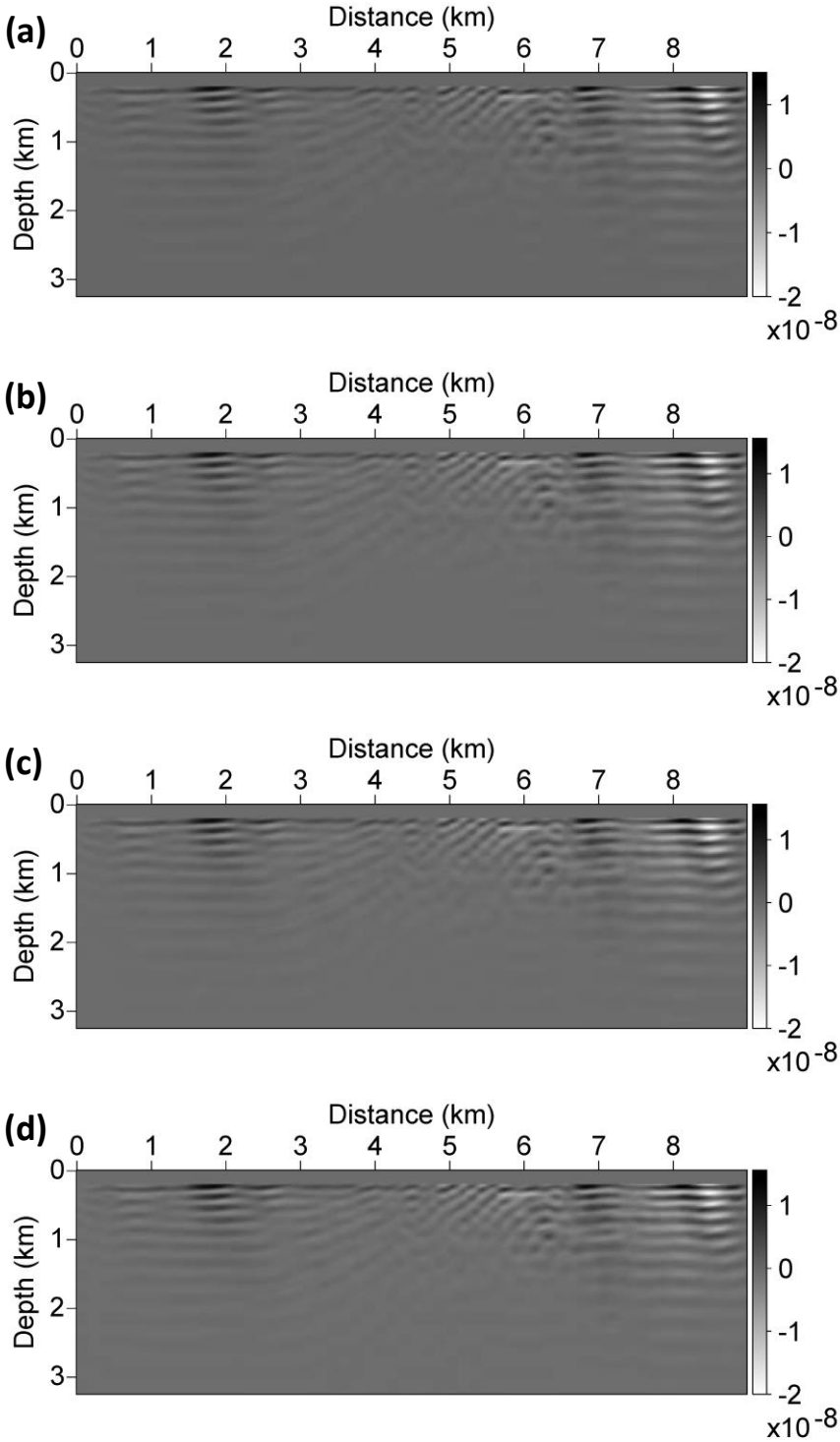

Figure 3.6 The matrix-vector product of the Hessian matrix and column vector $\mathbf{H}_i\mathbf{p}$ at 5.5 Hz obtained using the forward FDA method with intervals of (a) 0.1, (b) 0.01, (c) 0.001 and (d) 0.0001 for the Marmousi model.

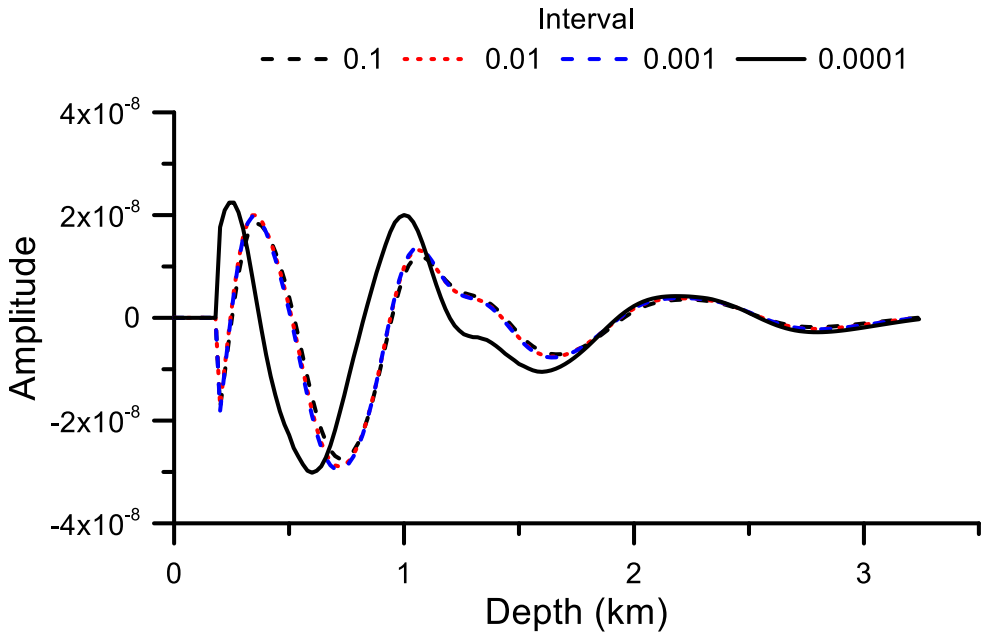

Figure 3.7 Depth profiles of $\mathbf{H}_i\mathbf{p}$ extracted at a distance of 6.9 km for 1.5 Hz. The forward FDA method is used to approximate $\mathbf{H}_i\mathbf{p}$ with intervals of 0.1 (black dashed line), 0.01 (red dotted line), 0.001 (blue dashed line) and 0.0001 (black solid line).

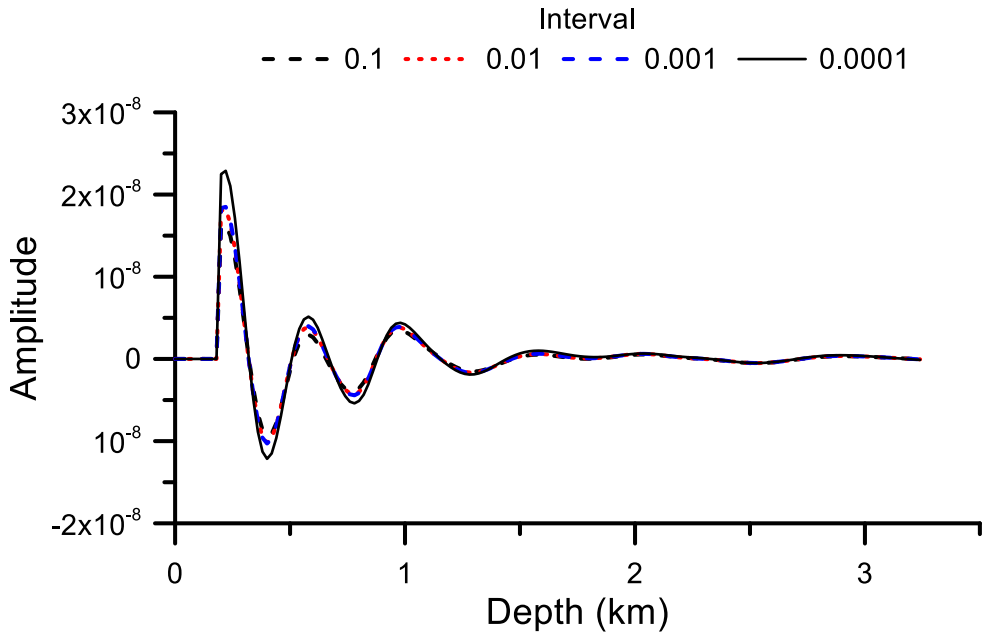

Figure 3.8 Depth profiles of $\mathbf{H}_i\mathbf{p}$ extracted at a distance of 6.9 km for 2.5 Hz. The forward FDA method is used to approximate $\mathbf{H}_i\mathbf{p}$ with intervals of 0.1 (black dashed line), 0.01 (red dotted line), 0.001 (blue dashed line) and 0.0001 (black solid line).

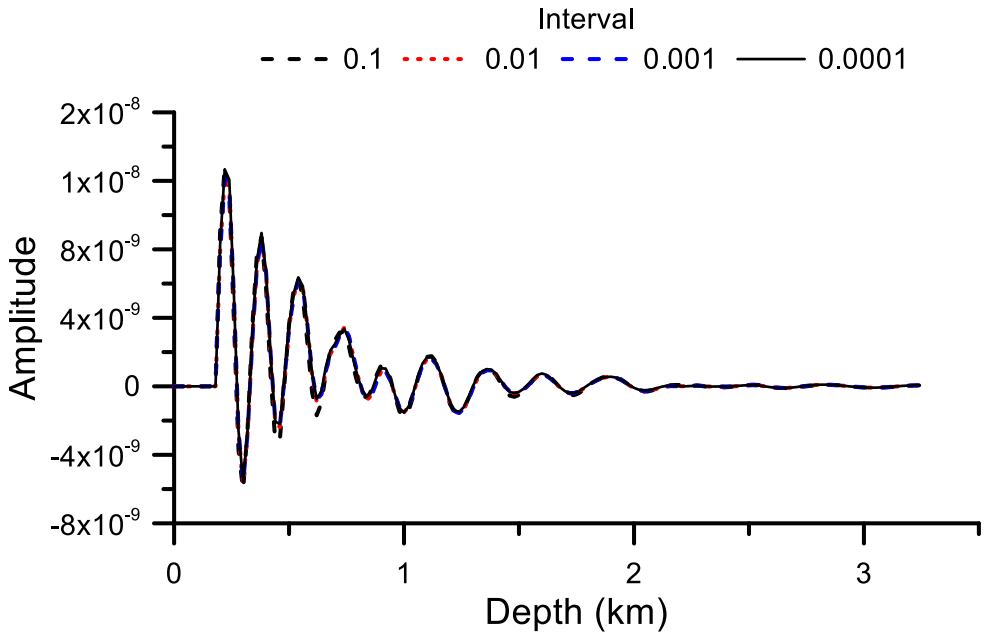

Figure 3.9 Depth profiles of $\mathbf{H}_i\mathbf{p}$ extracted at a distance of 6.9 km for 5.5 Hz. The forward FDA method is used to approximate $\mathbf{H}_i\mathbf{p}$ with intervals of 0.1 (black dashed line), 0.01 (red dotted line), 0.001 (blue dashed line) and 0.0001 (black solid line).

3.1.3 Analysis of model dependency

Numerical tests for the Marmousi model show that using too small interval ε (e.g., lower than 0.0001) can cause the failure of the $\mathbf{Hp}$ approximation because of round-off errors depending on the FDA method and frequency range. To avoid round-off errors, a sufficiently large interval is needed for the successful $\mathbf{Hp}$ approximation. Numerical tests for the Marmousi model show that large intervals, which are larger than 0.0001, approximate the Hessian approximation stably and accurately. However, a large interval like 0.1, 0.01 and 0.001 may fail to approximate $\mathbf{Hp}$ accurately because the approximation errors of the forward and central FDA methods depend on $O(\varepsilon)$ and $O(\varepsilon^2)$, respectively (equations 2.29 and 2.30).

In addition, the steepest-descent directions, which are needed to approximate $\mathbf{Hp}$, have the different amplitudes depending on the model parameters like true, initial and inverted models. In this section, numerical tests for the acoustic Overthrust model are provided to investigate the problem generated by model parameters. Table 3.2 shows the FWI settings. Figure 3.10 shows the true model for the acoustic Overthrust model and the initial model used for inversion. In the initial model, the velocity increases linearly from 3 km/s to 6.5 km/s. Consider that the interval, used to approximate $\mathbf{Hp}$ stably and accurately for the Marmousi model, can fail to approximate $\mathbf{Hp}$ for the acoustic Overthrust model. With the poorly or wrongly estimated $\mathbf{Hp}$ approximation, the search direction of model parameters, which is derived by solving the linear conjugate-gradient method, will fail to minimize the objective function. Figure 3.11 shows the search directions of P-wave velocity obtained at the first iteration for different approximation intervals. Search directions obtained using intervals of 0.1 and 0.01 (Figures 3.11a and 3.11b) show

reverse search directions compared with that obtained using an interval of 0.001 (Figure 3.11c). To investigate which search directions minimize the objective function, RMS error curves are compared in Figure 3.12. The RMS error curve increases or oscillates, which means that the HF optimization method with intervals of 0.1, 0.01, 0.001 and 0.0001 fails in solving the nonlinear problem because of the failure of $\mathbf{Hp}$ approximation.

Table 3.2 Inversion settings used for numerical tests of the acoustic Overthrust model.

Model size	No. of shot	No. of receiver	Shot interval	Recording time	Maximum Frequency	Minimum Frequency
9.6 km × 3.4 km	157	470	0.06 km	8 sec	10 Hz	0.25 Hz

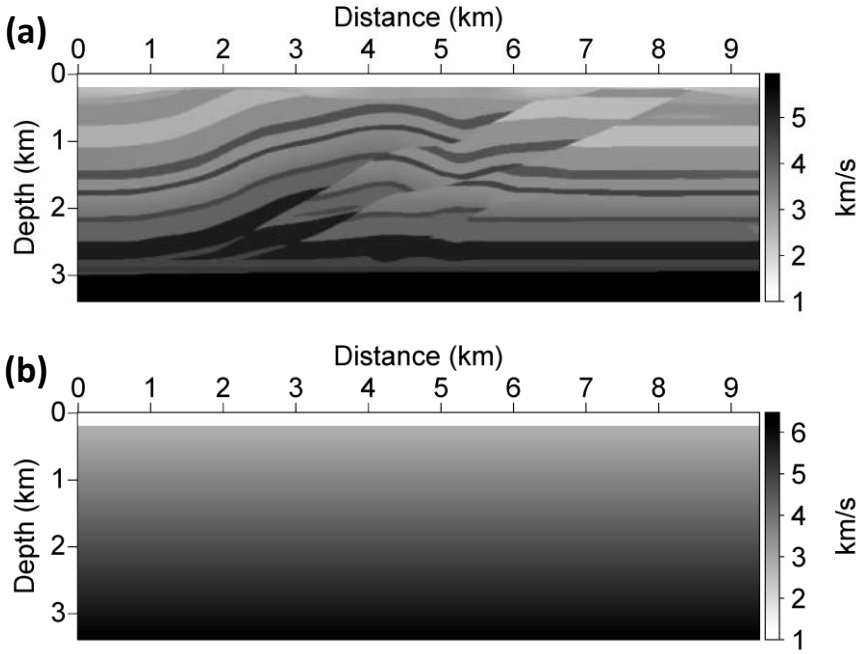

Figure 3.10 P-wave velocities of (a) the true acoustic Overthrust and (b) linearly-increasing initial models used to analyze model dependency of the **Hp** approximation using the forward FDA method.

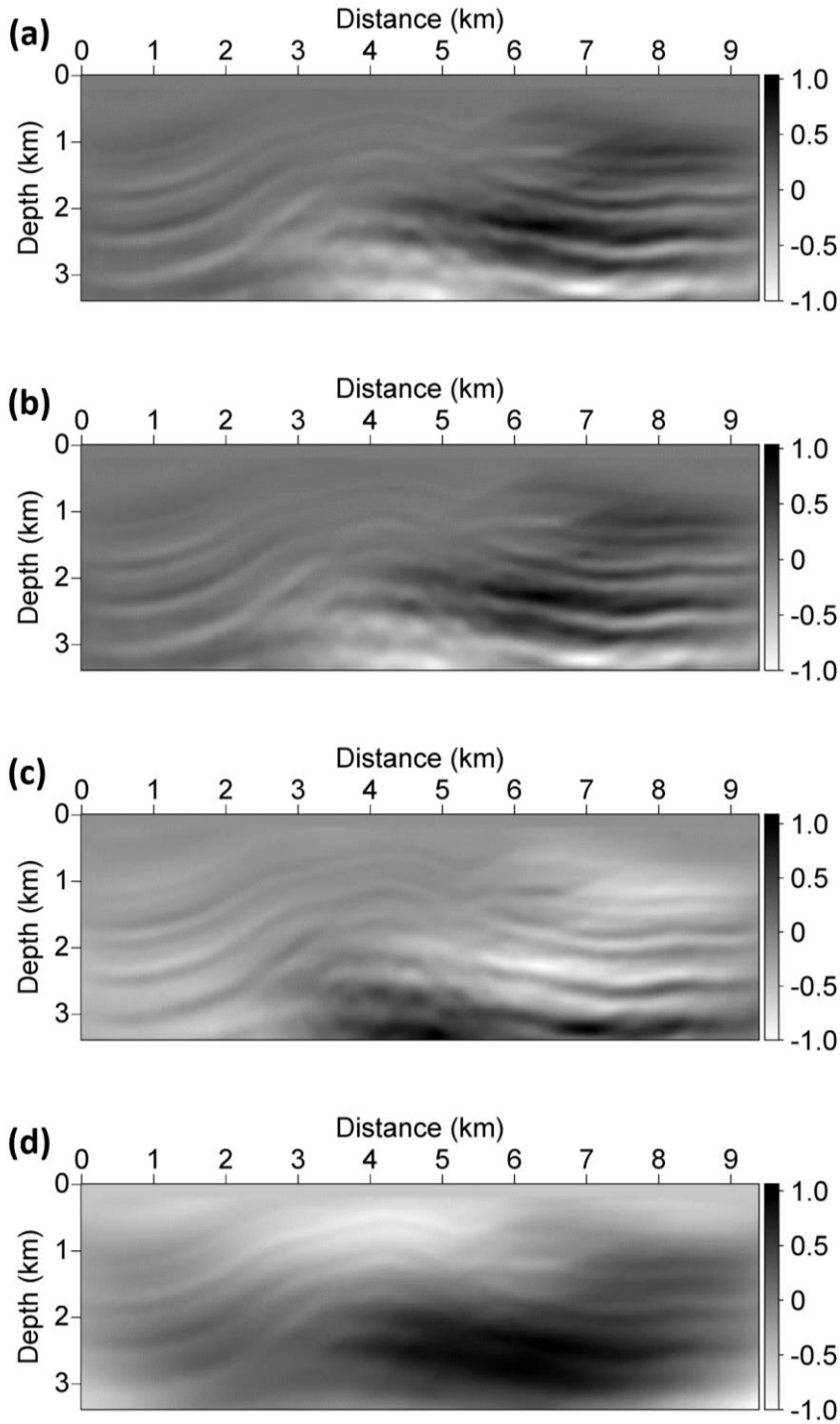

Figure 3.11 Comparisons of search directions obtained by the HF optimization method using forward FDA method with intervals of (a) 0.1, (b) 0.01, (c) 0.001 and (d) 0.0001 for the acoustic Overthrust model.

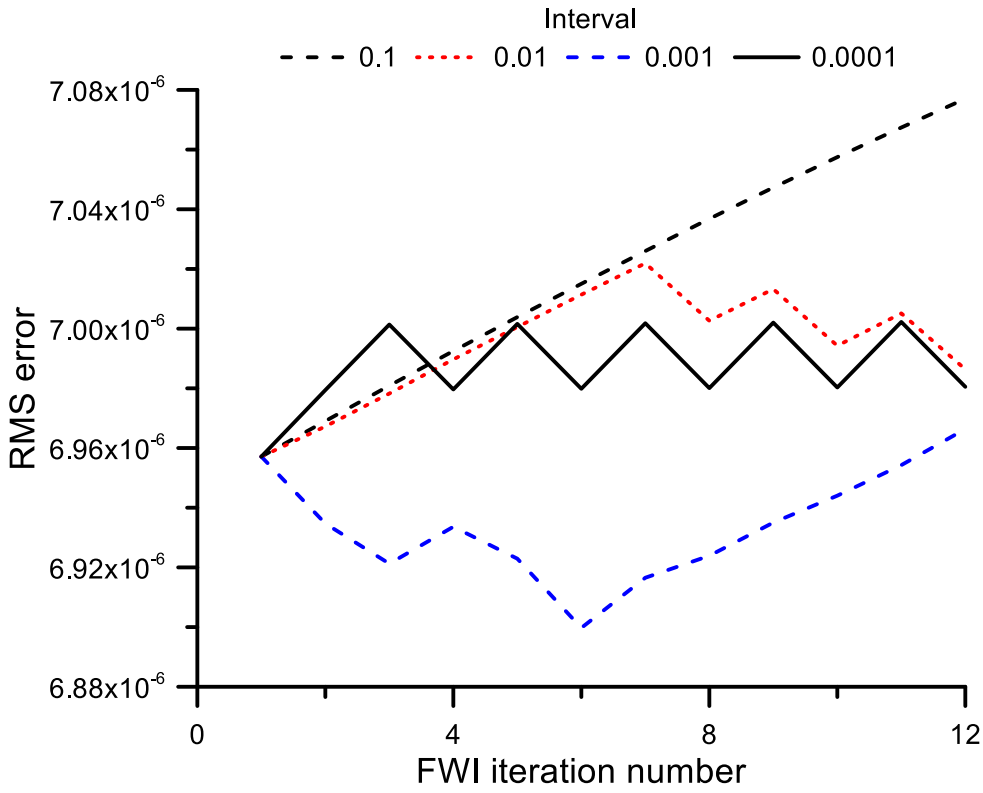

Figure 3.12 RMS error curves obtained by the HF optimization method using the forward FDA method with intervals of 0.1 (black dashed line), 0.01 (red dotted line), 0.001 (blue dashed line) and 0.0001 (black solid line).

Numerical tests for the Marmousi and acoustic Overthrust models show that the approximation interval ε which is used to approximate $\mathbf{Hp}$ is very sensitive to model parameters because the steepest-descent directions vary depending on model parameters. In the same way, during seismic FWI, the model vector $\mathbf{m}^i$ at the i^{th} FWI iteration changes as FWI iteration proceeds. To demonstrate this phenomenon, FWI using the HF optimization is performed for the acoustic Overthrust model assuming the approximation intervals ε to be 0.01, 0.001 and 0.0001. All inversion settings are the same as those in Table 3.2. The linearly-increasing initial model ranges from 3 km/s to 6 km/s. RMS error curves are plotted in Figure 3.13. The RMS error curve with an interval of 0.0001 converges at the 6^{th} iteration, although the others for intervals of 0.01 and 0.001 are still decreasing after the 6^{th} iteration. In other words, the HF optimization using an approximation interval of 0.0001 is stuck in local minimum. To escape from the local minimum, the approximation interval is changed from 0.0001 to 0.001 at the 6^{th} iteration. Figure 3.14 shows RMS error curves after changing the approximation interval. From Figure 3.14, we see that the HF optimization using changed approximation interval converges to solutions close to the global minimum just like the other cases using intervals of 0.01 and 0.001.

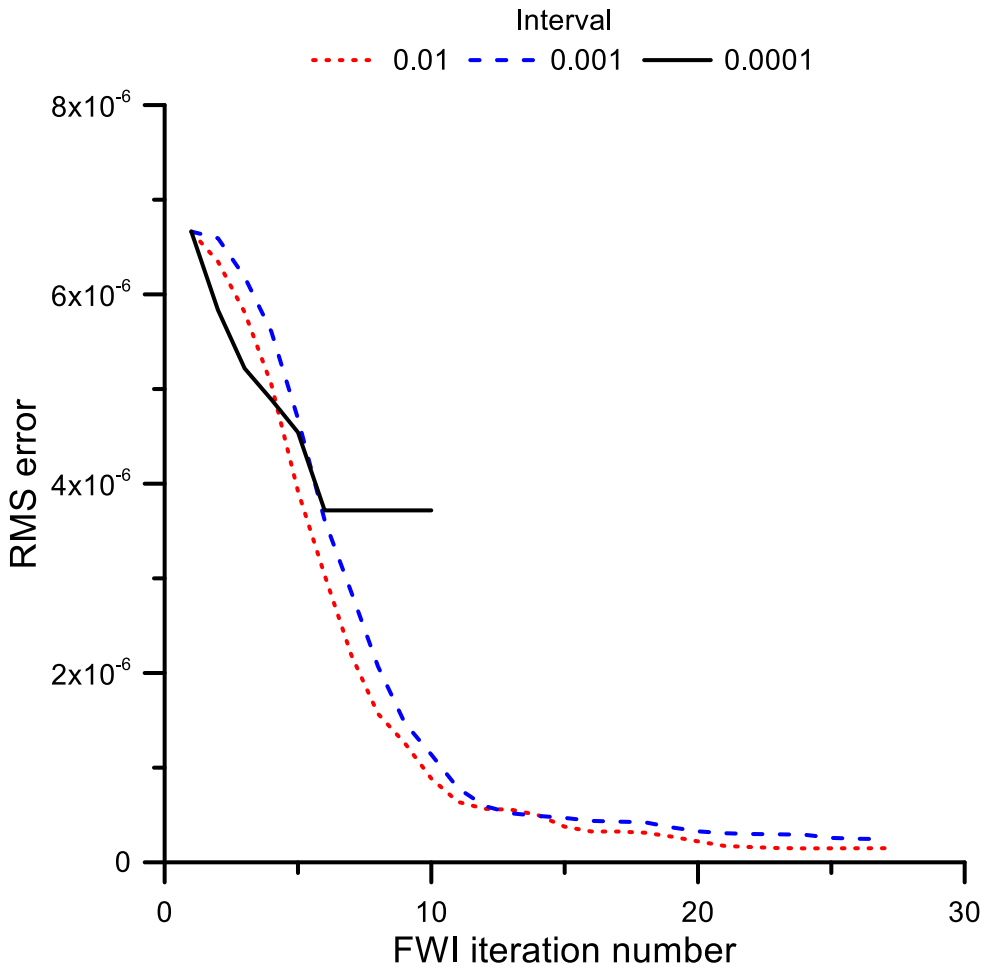

Figure 3.13 RMS error curves obtained by the HF optimization method using the forward FDA method with intervals of 0.01 (red dotted line), 0.001 (blue dashed line) and 0.0001 (black solid line).

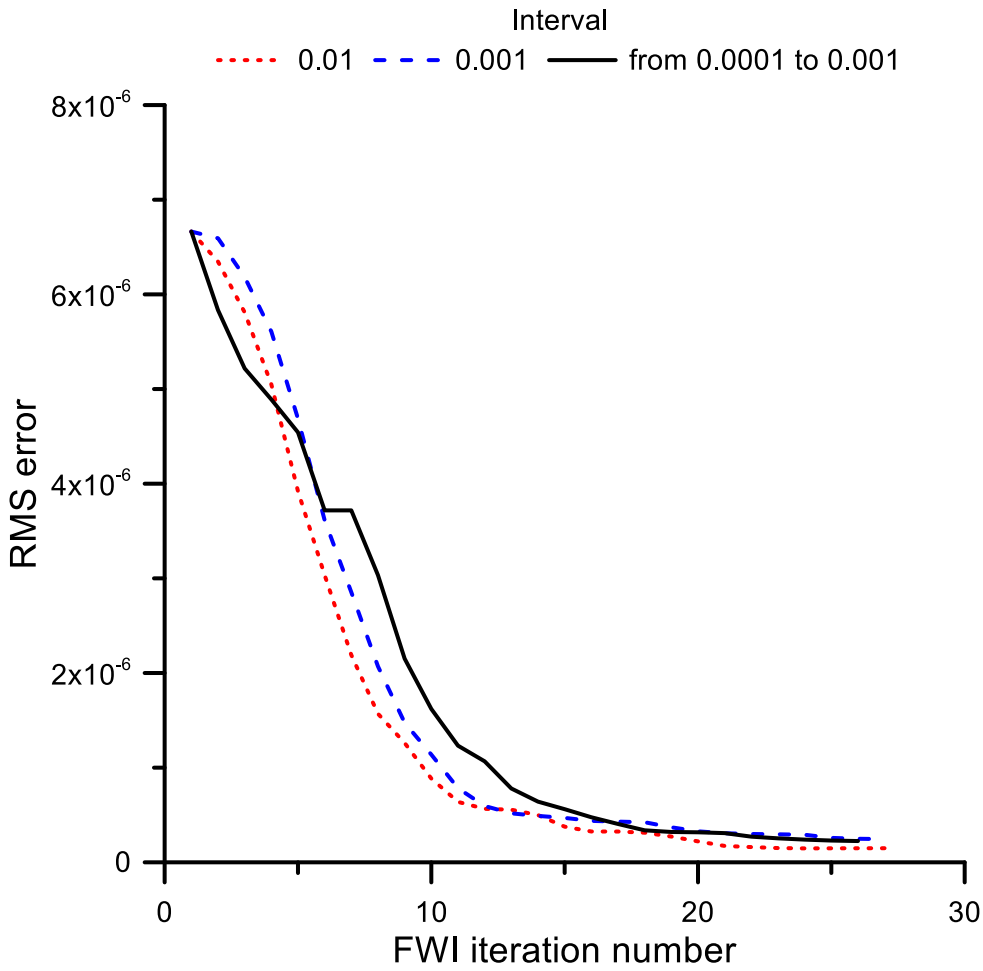

Figure 3.14 RMS error curves obtained by the HF optimization method using the forward FDA method with intervals of 0.01 (red dotted line) and 0.001 (blue dashed line) and the interval changing from 0.0001 to 0.001 (black solid line).

3.1.4 Limitation of the Hessian approximation

The HF optimization method does not require the exact information of the Hessian matrix itself or its inverse although it is a kind of the second-order optimization methods. Instead of using the exact Hessian matrix, the HF optimization method uses the approximation of the matrix-vector product of the Hessian matrix and column vector $\mathbf{Hp}$.

The $\mathbf{Hp}$ approximation is derived using the forward or the central FDA methods which require additional gradient calculations per a linear conjugate-gradient iteration as described in equations (2.29) and (2.30). From Figures 3.2a ~ 3.2c and 3.3a ~ 3.3c, it is noted that both the forward and central FDA methods yield similar approximations of $\mathbf{Hp}$. However, the forward FDA method fails to approximate $\mathbf{Hp}$ with intervals of 0.0001 and 0.00001, whereas the central FDA method can approximate $\mathbf{Hp}$ with an interval of 0.0001. In addition, using an interval of 0.00001 fails to approximate $\mathbf{Hp}$ in both methods because of the round-off errors. Although the central FDA method is more stable and accurate than the forward FDA method, the forward FDA method has been preferred to the central FDA method due to reduction of computational cost related with the number of forward modeling.

The first-order partial derivative of objective function, $\nabla E\left(\mathbf{m}_0 + \varepsilon\mathbf{p}\right)$, is affected by inversion settings like true and initial model parameters (Figures 3.11 and 3.12), and frequencies (Figures 3.4~3.6). To determine an appropriate approximation interval ε, the trial and error need to be applied. In other words, to approximate $\mathbf{Hp}$ accurately and stably, a number of numerical tests should be performed to find an appropriate approximation interval ε. In addition, an appropriate interval ε chosen to approximate the $\mathbf{Hp}$ accurately in the early

stage of FWI may not be appropriate for the later stage of FWI, and thus the convergence to the global minimum is not guaranteed as shown in Figure 3.13.

3.2 The improved Hessian-Free optimization

As mentioned in section 3.1.4, an appropriate approximation interval ε determines needs to be chosen for a success of the HF optimization method in reaching the global minimum. In this study, to resolve the problems caused by using the FDA method, a new approximation method is proposed to approximate the matrix-vector product of the Hessian matrix and column vector, $\mathbf{Hp}$, without using interval ε. The new method does not require a number of numerical tests, which are required to determine an appropriate interval ε for the FDA method. In section 3.2.1, the new approximation method will be introduced. The derivation of the new approximation method is based on the limit of a function. After that, $\mathbf{Hp}$ approximated by the FDA method and the new approximation method will be compared to investigate the stability and accuracy of the improved HF optimization method.

3.2.1 Theory of the improved Hessian-Free optimization

The derivation of the improved HF optimization method is based on the limit of a function. Starting from equation (2.30) and considering the definition of the partial derivatives, equation (2.30) can be rewritten as follows:

$$\mathbf{H}(\mathbf{m}_0)\mathbf{p} = \nabla^2\mathbf{E}(\mathbf{m}_0)\mathbf{p} = \lim_{\varepsilon\to 0}\frac{\nabla E(\mathbf{m}_0+\varepsilon\mathbf{p}) - \nabla E(\mathbf{m}_0-\varepsilon\mathbf{p})}{2\varepsilon}, \qquad (3.4)$$

where the error term $O(\varepsilon^2)$ is missing because the approximation interval ε goes to 0 by the definition of the limit. Note that the approximation interval ε, which is discussed in the previous section 3.1, is not 0 but a finite small real value and the choice of the approximation interval ε requires sophisticated and careful efforts depending on inversion settings. The new approximation method is designed to be independent of the approximation interval by making the approximation intervals ε cancelled in numerator and denominator of equation (3.4).

The steepest-descent directions can be efficiently calculated using the backpropagation technique (Pratt et al., 1998) as described in equation (2.9). With the virtual source, equation (3.4) can be rewritten as:

$$\mathbf{H}(\mathbf{m}_0)\mathbf{p} = \lim_{\varepsilon\to 0}\left[\begin{array}{c}\dfrac{\mathbf{v}(\mathbf{m}_0+\varepsilon\mathbf{p})^t\,\mathbf{S}(\mathbf{m}_0+\varepsilon\mathbf{p})^{-1}\left(\mathbf{L}\mathbf{u}(\mathbf{m}_0+\varepsilon\mathbf{p})-\mathbf{d}\right)^*}{2\varepsilon} \\ -\dfrac{\mathbf{v}(\mathbf{m}_0-\varepsilon\mathbf{p})^t\,\mathbf{S}(\mathbf{m}_0-\varepsilon\mathbf{p})^{-1}\left(\mathbf{L}\mathbf{u}(\mathbf{m}_0-\varepsilon\mathbf{p})-\mathbf{d}\right)^*}{2\varepsilon}\end{array}\right], \qquad (3.5)$$

where the matrix $\mathbf{S}(\mathbf{m}_0 \pm \varepsilon\mathbf{p})$ is the modeling operator and $\mathbf{u}(\mathbf{m}_0 \pm \varepsilon\mathbf{p})$

indicates the modeled data with the perturbed P-wave velocity vector $\mathbf{m}_0 \pm \varepsilon\mathbf{p}$, respectively. Equation (2.3) can also be rewritten with respect to the model vector $\mathbf{m}_0$ as follows:

$$\mathbf{S}\left(\mathbf{m}_0\right)\mathbf{u}(\mathbf{m}_0) = \mathbf{S}_0\mathbf{u}_0 = \mathbf{f}\,, \tag{3.6}$$

where $\mathbf{S}_0$ indicates the modeling operator and $\mathbf{u}_0$ denotes the modeled data with $\mathbf{m}_0$ (e.g., P-wave velocity). In the same manner, the forward modeling equation can be expressed with perturbed model parameter $\mathbf{m}_0 \pm \varepsilon\mathbf{p}$ as:

$$\mathbf{S}\left(\mathbf{m}_0 \pm \varepsilon\mathbf{p}\right)\mathbf{u}\left(\mathbf{m}_0 \pm \varepsilon\mathbf{p}\right) = \mathbf{f}\,. \tag{3.7}$$

By defining the vector $\Delta\mathbf{u}^{\pm}$ and matrix $\Delta\mathbf{S}^{\pm}$ as the perturbed modeled wavefields and the perturbed modeling operator, respectively, equation (3.7) can be rewritten as follows:

$$\left(\mathbf{S}_0 + \Delta\mathbf{S}^{\pm}\right)\left(\mathbf{u}_0 + \Delta\mathbf{u}^{\pm}\right) = \mathbf{f} \tag{3.8}$$

with relationships of $\mathbf{S}_0 = \mathbf{S}\left(\mathbf{m}_0\right)$, $\mathbf{u}_0 = \mathbf{u}\left(\mathbf{m}_0\right)$, $\mathbf{S}\left(\mathbf{m}_0 \pm \varepsilon\mathbf{p}\right) = \mathbf{S}_0 + \Delta\mathbf{S}^{\pm}$ and $\mathbf{u}\left(\mathbf{m}_0 \pm \varepsilon\mathbf{p}\right) = \mathbf{u}_0 + \Delta\mathbf{u}^{\pm}$. From the forward modeling equations (3.6) and (3.8), a new relationship can be introduced as follows:

$$\mathbf{S}_0\Delta\mathbf{u}^{\pm} = -\Delta\mathbf{S}^{\pm}\mathbf{u}_0\,, \tag{3.9}$$

where $\Delta\mathbf{S}^{\pm}\Delta\mathbf{u}^{\pm}$ is neglected. The perturbed modeling operator and the perturbed

modeled wavefields can be expressed as:

$$\Delta\mathbf{S}^{\pm} = \mathbf{S}\left(\mathbf{m}_0 \pm \varepsilon\mathbf{p}\right) - \mathbf{S}_0 =$$

$$\begin{bmatrix} -\frac{\omega^2}{\left(m_0 \pm \varepsilon p\right)_1^2} + \frac{\omega^2}{\left(m_0\right)_1^2} & & & & \\ & \ddots & & & \\ & & -\frac{\omega^2}{\left(m_0 \pm \varepsilon p\right)_k^2} + \frac{\omega^2}{\left(m_0\right)_k^2} & & \\ & & & \ddots & \\ & & & & -\frac{\omega^2}{\left(m_0 \pm \varepsilon p\right)_{nxz}^2} + \frac{\omega^2}{\left(m_0\right)_{nxz}^2} \end{bmatrix} \tag{3.10}$$

$$\Delta\mathbf{u}^{\pm} = \mathbf{S}_0^{-1}\left(-\Delta\mathbf{S}^{\pm}\mathbf{u}_0\right), \tag{3.11}$$

where $\Delta\mathbf{S}^{\pm}$ is a diagonal matrix. The k[th] diagonal element of $\Delta\mathbf{S}^{\pm}$ in equation (3.10) can be expressed as follows:

$$\Delta\mathbf{S}_k^{\pm} = \omega^2\left[\frac{\pm 2\left(\varepsilon m_0 p\right)_k + \left(\varepsilon p\right)_k^2}{\left(m_0^2 \pm \varepsilon p m_0\right)_k^2}\right]. \tag{3.12}$$

In the same manner, the perturbed back-propagated wavefield vector $\Delta\mathbf{b}$, which is expressed in equation (2.9), can also be derived by

$$\mathbf{b}_0 = \mathbf{S}_0^{-1}\left(\mathbf{L}\mathbf{u}_0 - \mathbf{d}\right)^*, \tag{3.13}$$

$$\mathbf{S}_0\mathbf{b}_0 = \left(\mathbf{L}\mathbf{u}_0 - \mathbf{d}\right)^*, \tag{3.14}$$

where $\mathbf{b}_0$ indicates the back-propagated wavefields with the model parameter $\mathbf{m}_0$ and the conjugated residual vector $\left(\mathbf{L}\mathbf{u}_0 - \mathbf{d}\right)^*$ behaves similarly just like the seismic source vector $\mathbf{f}$ in equation (3.6). With the perturbed modeling operator $\Delta\mathbf{S}^{\pm}$, the perturbed back-propagated wavefields $\Delta\mathbf{b}^{\pm}$ can be expressed as:

$$\left(\mathbf{S}_0 + \Delta\mathbf{S}^{\pm}\right)\left(\mathbf{b}_0 + \Delta\mathbf{b}^{\pm}\right) = \left(\mathbf{L}\left(\mathbf{u}_0 + \Delta\mathbf{u}^{\pm}\right) - \mathbf{d}\right)^*. \tag{3.15}$$

By substituting equation (3.14) into equation (3.15) and neglecting $\Delta\mathbf{S}^{\pm}\Delta\mathbf{b}^{\pm}$ just like equation (3.9), the perturbed back-propagated wavefields can be written as follows:

$$\mathbf{S}_0\Delta\mathbf{b}^{\pm} + \Delta\mathbf{S}^{\pm}\mathbf{b}_0 = \left(\mathbf{L}\Delta\mathbf{u}^{\pm}\right)^*, \tag{3.16}$$

$$\Delta\mathbf{b}^{\pm} = \mathbf{S}_0^{-1}\left(-\Delta\mathbf{S}^{\pm}\mathbf{b}_0 + \left(\mathbf{L}\Delta\mathbf{u}^{\pm}\right)^*\right). \tag{3.17}$$

Consequently, deriving the perturbed modeled wavefields and the perturbed back-propagated wavefields, which are expressed as equations (3.11) and (3.17), requires computation of $\mathbf{S}_0^{-1}$ with the model parameter $\mathbf{m}_0$ twice. In other words, construction and factorization of the new modeling operator $\left(\mathbf{S}_0 + \Delta\mathbf{S}^{\pm}\right)$ with the model parameter $\mathbf{m}_0 \pm \varepsilon\mathbf{p}$ are not needed in iterations for the linear conjugate-gradient method. Substituting equation (3.15) into equation (3.5) yields

$$\mathbf{H}\left(\mathbf{m}_0\right)\mathbf{p} = \lim_{\varepsilon\to 0}\left[\frac{\mathbf{v}\left(\mathbf{m}_0 + \varepsilon\mathbf{p}\right)^t\left(\mathbf{b}_0 + \Delta\mathbf{b}^{+}\right) - \mathbf{v}\left(\mathbf{m}_0 - \varepsilon\mathbf{p}\right)^t\left(\mathbf{b}_0 + \Delta\mathbf{b}^{-}\right)}{2\varepsilon}\right]. \tag{3.18}$$

The virtual source $\mathbf{v}$ (defined in equation 2.8) with the small perturbed model parameter $\mathbf{m}_0 \pm \varepsilon \mathbf{p}$ in equation (3.18) can be derived using the perturbed modeled wavefields as follows:

$$\mathbf{v}_k\left(\mathbf{m}_0 \pm \varepsilon \mathbf{p}\right) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ -\dfrac{2\omega^2\left(u_0 + \Delta u^{\pm}\right)_k}{\left(m_0 \pm \varepsilon p\right)_k^3} \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{3.19}$$

By substituting equation (3.19) into (3.18), the kth element in equation (3.18) can be rewritten as:

$$\left[\mathbf{H}\left(\mathbf{m}_0\right)\mathbf{p}\right]_k = \lim_{\varepsilon \to 0} \left[\begin{array}{l} \dfrac{-2\omega^2\left(u_0 + \Delta u^{+}\right)_k\left(b_0 + \Delta b^{+}\right)_k}{2\varepsilon\left(m_0 + \varepsilon p\right)_k^3} \\ + \dfrac{2\omega^2\left(u_0 + \Delta u^{-}\right)_k\left(b_0 + \Delta b^{-}\right)_k}{2\varepsilon\left(m_0 - \varepsilon p\right)_k^3} \end{array} \right]$$

$$= \omega^2 \lim_{\varepsilon \to 0} \left[\begin{array}{l} \dfrac{\left(u_0 b_0\right)_k\left(-\left(m_0 - \varepsilon p\right)_k^3 + \left(m_0 + \varepsilon p\right)_k^3\right)}{\varepsilon\left(m_0 + \varepsilon p\right)_k^3\left(m_0 - \varepsilon p\right)_k^3} \\ + \dfrac{\left(\Delta u^{+} b_0\right)_k\left(-\left(m_0 - \varepsilon p\right)_k^3\right) + \left(\Delta u^{-} b_0\right)_k\left(\left(m_0 + \varepsilon p\right)_k^3\right)}{\varepsilon\left(m_0 + \varepsilon p\right)_k^3\left(m_0 - \varepsilon p\right)_k^3} \\ + \dfrac{\left(u_0 \Delta b^{+}\right)_k\left(-\left(m_0 - \varepsilon p\right)_k^3\right) + \left(u_0 \Delta b^{-}\right)_k\left(\left(m_0 + \varepsilon p\right)_k^3\right)}{\varepsilon\left(m_0 + \varepsilon p\right)_k^3\left(m_0 - \varepsilon p\right)_k^3} \end{array} \right]. \tag{3.20}$$

The first term in the right hand side of equation (3.20) becomes

$$\begin{aligned}&\omega^2\left(u_0b_0\right)_k\lim_{\varepsilon\to0}\left[\frac{\left(-\left(m_0-\varepsilon p\right)_k^3+\left(m_0+\varepsilon p\right)_k^3\right)}{\varepsilon\left(m_0+\varepsilon p\right)_k^3\left(m_0-\varepsilon p\right)_k^3}\right].\\&=\omega^2\left(u_0b_0\right)_k\lim_{\varepsilon\to0}\left[\frac{2\varepsilon p_k\left(3m_0^2+\varepsilon^2p^2\right)_k}{\varepsilon\left(m_0+\varepsilon p\right)_k^3\left(m_0-\varepsilon p\right)_k^3}\right]\end{aligned}\qquad(3.21)$$

The second term in the right hand side of equation (3.20) reduces to

$$\begin{aligned}&\omega^2\left(b_0\right)_k\lim_{\varepsilon\to0}\left[\frac{\left(\Delta u^+\right)_k\left(-\left(m_0-\varepsilon p\right)_k^3\right)+\left(\Delta u^-\right)_k\left(\left(m_0+\varepsilon p\right)_k^3\right)}{\varepsilon\left(m_0+\varepsilon p\right)_k^3\left(m_0-\varepsilon p\right)_k^3}\right]\\&=\omega^2\left(b_0\right)_k\lim_{\varepsilon\to0}\left[\begin{array}{l}-\dfrac{\left(\Delta u^+-\Delta u^-\right)_k\left(m_0^3+3m_0\varepsilon^2p^2\right)_k}{\varepsilon\left(m_0+\varepsilon p\right)_k^3\left(m_0-\varepsilon p\right)_k^3}\\+\dfrac{\left(\Delta u^++\Delta u^-\right)_k\left(3m_0^2\varepsilon p+\varepsilon^3p^3\right)_k}{\varepsilon\left(m_0+\varepsilon p\right)_k^3\left(m_0-\varepsilon p\right)_k^3}\end{array}\right].\end{aligned}\qquad(3.22)$$

The third term in the right hand side of equation (3.20) is expressed as follows:

$$\begin{aligned}&\omega^2\left(u_0\right)_k\lim_{\varepsilon\to0}\left[\frac{\left(\Delta b^+\right)_k\left(-\left(m_0-\varepsilon p\right)_k^3\right)+\left(\Delta b^-\right)_k\left(\left(m_0+\varepsilon p\right)_k^3\right)}{\varepsilon\left(m_0+\varepsilon p\right)_k^3\left(m_0-\varepsilon p\right)_k^3}\right]\\&=\omega^2\left(u_0\right)_k\lim_{\varepsilon\to0}\left[\begin{array}{l}-\dfrac{\left(\Delta b^+-\Delta b^-\right)_k\left(m_0^3+3m_0\varepsilon^2p^2\right)_k}{\varepsilon\left(m_0+\varepsilon p\right)_k^3\left(m_0-\varepsilon p\right)_k^3}\\+\dfrac{\left(\Delta b^++\Delta b^-\right)_k\left(3m_0^2\varepsilon p+\varepsilon^3p^3\right)_k}{\varepsilon\left(m_0+\varepsilon p\right)_k^3\left(m_0-\varepsilon p\right)_k^3}\end{array}\right].\end{aligned}\qquad(3.23)$$

In equation (3.21), the approximation intervals are cancelled out and, by the definition of the limit, the remaining ε goes to zero as follows:

$$\begin{aligned}&\omega^2\left(u_0 b_0\right)_k \lim_{\varepsilon\to 0}\left[\frac{2\cancel{\varepsilon}\, p_k\left(3m_0^2+\varepsilon^2 p^2\right)_k}{\cancel{\varepsilon}\left(m_0+\varepsilon p\right)_k^3\left(m_0-\varepsilon p\right)_k^3}\right].\\&=6\omega^2\left(u_0 b_0\frac{p}{m_0^4}\right)_k\end{aligned} \tag{3.24}$$

In equations (3.22) and (3.23), ε does not appear in the numerator unlike in equation (3.21). However, the perturbed modeled wavefields $\Delta\mathbf{u}$ and the perturbed back-propagated wavefields $\Delta\mathbf{b}$ in equations (3.22) and (3.23) have ε in themselves according to equations (3.11), (3.12) and (3.17). After dividing equation (3.12) by ε and taking the limit, equations (3.25), (3.26) and (3.27) can be written as follows:

$$\lim_{\varepsilon\to 0}\frac{\Delta\mathbf{S}_k^{\pm}}{\varepsilon}=\lim_{\varepsilon\to 0}\omega^2\left[\frac{\pm 2\cancel{\varepsilon}\left(m_0 p\right)_k+\cancel{\varepsilon^2}\, p_k^2}{\cancel{\varepsilon}\left(m_0^2\pm\varepsilon p m_0\right)_k^2}\right]=\pm 2\omega^2\left(\frac{p}{m_0^3}\right)_k, \tag{3.25}$$

$$\lim_{\varepsilon\to 0}\frac{\Delta\mathbf{u}^{\pm}}{\varepsilon}=\lim_{\varepsilon\to 0}\mathbf{S}_0^{-1}\left(\frac{-\Delta\mathbf{S}^{\pm}}{\varepsilon}\mathbf{u}_0\right)=\mathbf{S}_0^{-1}\left[\lim_{\varepsilon\to 0}\left(\frac{-\Delta\mathbf{S}^{\pm}}{\varepsilon}\right)\right]\mathbf{u}_0$$

$$=\mathbf{S}_0^{-1}\left[\begin{bmatrix}\mp 2\omega^2\left(\frac{p}{m_0^3}\right)_1 & & & & \\ & \ddots & & & \\ & & \mp 2\omega^2\left(\frac{p}{m_0^3}\right)_k & & \\ & & & \ddots & \\ & & & & \mp 2\omega^2\left(\frac{p}{m_0^3}\right)_{nxz}\end{bmatrix}\mathbf{u}_0\right], \quad (3.26)$$

$$\lim_{\varepsilon\to 0}\frac{\Delta\mathbf{b}^{\pm}}{\varepsilon}=\lim_{\varepsilon\to 0}\mathbf{S}_0^{-1}\left(\frac{\left(\mathbf{L}\Delta\mathbf{u}^{\pm}\right)^*}{\varepsilon}-\frac{\Delta\mathbf{S}^{\pm}}{\varepsilon}\mathbf{b}_0\right)$$

$$=\mathbf{S}_0^{-1}\left[\mathbf{L}\lim_{\varepsilon\to 0}\left(\frac{\left(\Delta\mathbf{u}^{\pm}\right)^*}{\varepsilon}\right)+\lim_{\varepsilon\to 0}\left(-\frac{\Delta\mathbf{S}^{\pm}}{\varepsilon}\right)\mathbf{b}_0\right]$$

$$=\mathbf{S}_0^{-1}\mathbf{L}\left[\lim_{\varepsilon\to 0}\frac{\left(\Delta\mathbf{u}^{\pm}\right)^*}{\varepsilon}\right]+$$

$$\mathbf{S}_0^{-1}\left[\begin{bmatrix}\mp 2\omega^2\left(\frac{p}{m_0^3}\right)_1 & & & & \\ & \ddots & & & \\ & & \mp 2\omega^2\left(\frac{p}{m_0^3}\right)_k & & \\ & & & \ddots & \\ & & & & \mp 2\omega^2\left(\frac{p}{m_0^3}\right)_{nxz}\end{bmatrix}\mathbf{b}_0\right]. \quad (3.27)$$

Because equations (3.22) and (3.23) contain $\Delta\mathbf{b}$ and $\Delta\mathbf{u}$ terms in themselves respectively, ε can be cancelled out as shown in equations (3.25), (3.26) and

(3.27). Using equations (3.26) and (3.27), equations (3.22) and (3.23) can be rewritten as follows:

$$\begin{aligned}&\omega^2\left(b_0\right)_k \lim_{\varepsilon\to 0}\left[-\frac{\left(\Delta u^+ - \Delta u^-\right)_k}{\varepsilon}\frac{\left(m_0^3 + 3m_0\varepsilon^2 p^2\right)_k}{\left(m_0+\varepsilon p\right)_k^3\left(m_0-\varepsilon p\right)_k^3}\right],\\ &=\omega^2\left(\frac{b_0}{m_0^3}\right)_k \lim_{\varepsilon\to 0}\left[-\frac{\left(\Delta u^+ - \Delta u^-\right)_k}{\varepsilon}\right]\end{aligned} \tag{3.28}$$

$$\begin{aligned}&\omega^2\left(u_0\right)_k \lim_{\varepsilon\to 0}\left[-\frac{\left(\Delta b^+ - \Delta b^-\right)_k}{\varepsilon}\frac{\left(m_0^3 + 3m_0\varepsilon^2 p^2\right)_k}{\left(m_0+\varepsilon p\right)_k^3\left(m_0-\varepsilon p\right)_k^3}\right],\\ &=\omega^2\left(\frac{u_0}{m_0^3}\right)_k \lim_{\varepsilon\to 0}\left[-\frac{\left(\Delta b^+ - \Delta b^-\right)_k}{\varepsilon}\right]\end{aligned} \tag{3.29}$$

Consequently, substituting equations (3.24), (3.28) and (3.29) into equation (3.20), the new method for the $\mathbf{Hp}$ approximation can be written as follows:

$$\begin{aligned}\left[\mathbf{H}\left(\mathbf{m}_0\right)\mathbf{p}\right]_k &= 6\omega^2\left(u_0 b_0\frac{p}{m_0^4}\right)_k\\ &+\omega^2\left(\frac{b_0}{m_0^3}\right)_k \lim_{\varepsilon\to 0}\left[-\frac{\left(\Delta u^+ - \Delta u^-\right)_k}{\varepsilon}\right],\\ &+\omega^2\left(\frac{u_0}{m_0^3}\right)_k \lim_{\varepsilon\to 0}\left[-\frac{\left(\Delta b^+ - \Delta b^-\right)_k}{\varepsilon}\right]\end{aligned} \tag{3.30}$$

where the second and the third terms in the right hand side of equation (3.30) can be derived using equations (3.26) and (3.27), respectively. From equation (3.30), it is noted that the new approximation of $\mathbf{Hp}$ is independent of the approximation interval ε.

3.2.2 Demonstration of the improved Hessian-Free optimization

To demonstrate the improved HF optimization method, numerical tests are performed with the same inversion settings as in Section 3.1. Figure 3.15 shows results of the **Hp** approximation obtained by the improved HF optimization method for the Marmousi model. For comparison, results obtained by the conventional optimization method (already presented in Figure 3.2) are also displayed in Figure 3.15. Results obtained by the improved HF optimization method look similar to those obtained by the conventional HF optimization method for intervals of 0.1, 0.01 and 0.001. To compare the accuracy of approximations more clearly, comparisons of depth profile for **Hp** are plotted in Figure 3.16. The depth profile obtained by the conventional HF optimization method using the interval of 0.1 shows slightly different values compared to the others because of the approximation error.

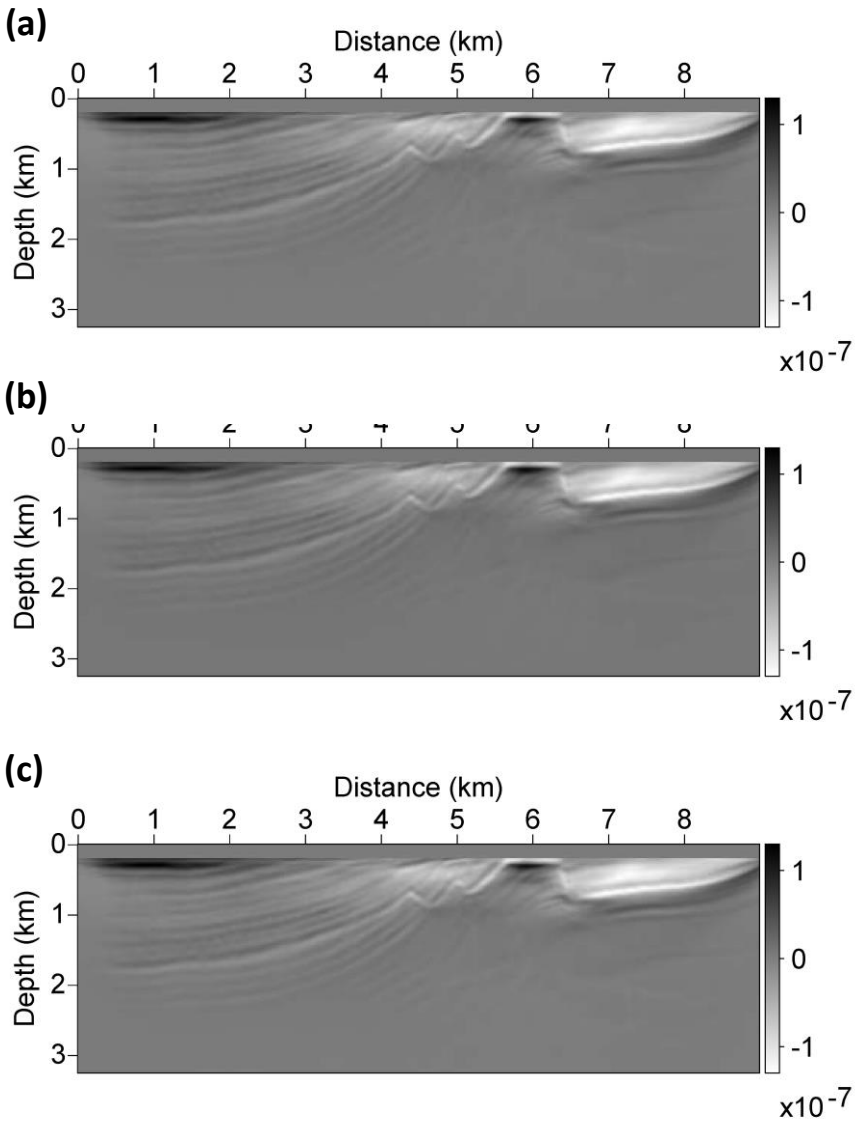

Figure 3.15 The matrix-vector product of the Hessian matrix and column vector **Hp** obtained by (a) the improved HF method and by the conventional HF method with intervals of (b) 0.1, (c) 0.01, (d) 0.001, (e) 0.0001 and (f) 0.00001 for the Marmousi model.

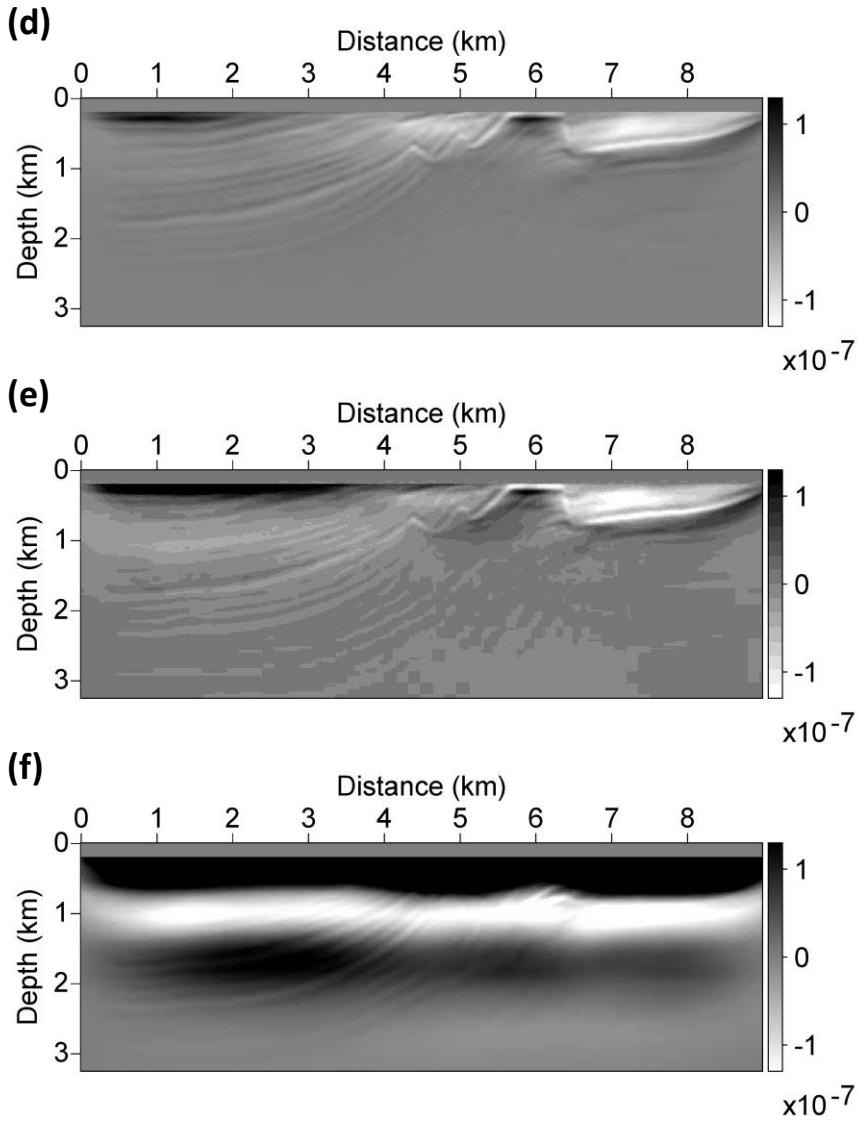

Figure 3.15 (Continued)

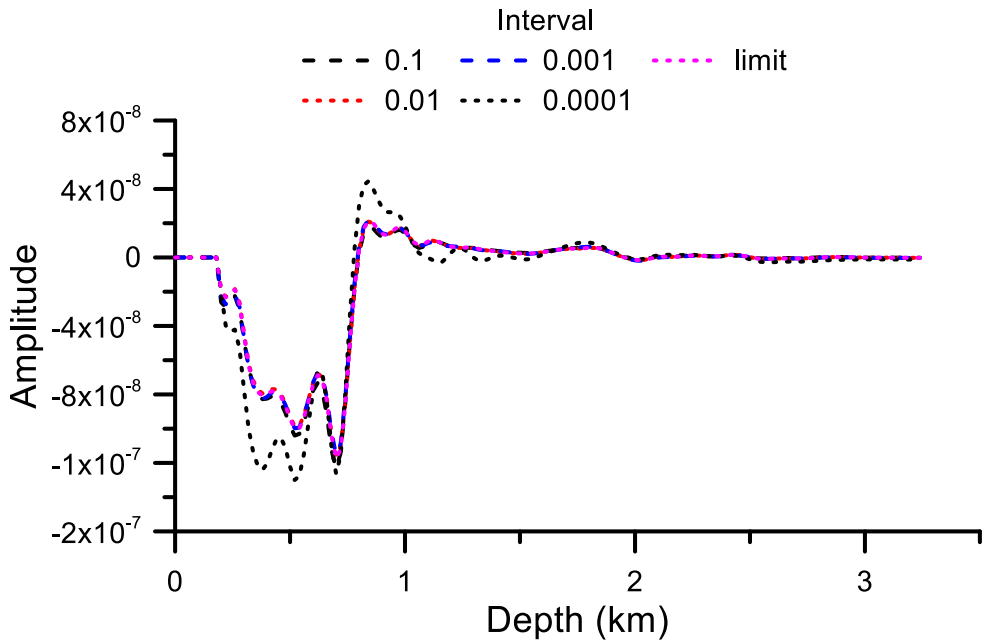

Figure 3.16 Depth profiles of **Hp** approximation obtained at a distance of 6.9 km by the improved HF method using the limit of a function (purple dotted line) and by the conventional HF method using the forward FDA method with intervals of 0.1 (black dashed line), 0.01 (red dotted line), 0.001 (blue dashed line) and 0.0001 (black dotted line).

As discussed in section 3.1.2, the accuracy and stability of $\mathbf{Hp}$ approximation depend on frequency, specifically low-frequency components. To investigate the accuracy and stability of $\mathbf{H}_i\mathbf{p}$ approximation obtained by the improved HF optimization method, numerical tests are performed with the same inversion settings as in section 3.1.2. Numerical results of $\mathbf{H}_i\mathbf{p}$ approximations obtained by the improved HF optimization method at 1.5, 2.5 and 5.5 Hz are shown in Figures 3.17(d), 3.18(d) and 3.19(d), respectively. $\mathbf{H}_i\mathbf{p}$ approximation obtained by the improved HF optimization method yields similar results to those obtained by the conventional method. To compare more precisely, depth profiles of $\mathbf{H}_i\mathbf{p}$ approximation extracted at a distance of 6.9 km are plotted in Figures 3.20, 3.21 and 3.23. $\mathbf{H}_i\mathbf{p}$ approximation results obtained by the improved HF optimization method match well with those obtained by the conventional HF optimization method with intervals of 0.1, 0.01 and 0.001. However, $\mathbf{H}_i\mathbf{p}$ approximation results obtained by the conventional HF optimization method with an interval of 0.1 show small approximation errors.

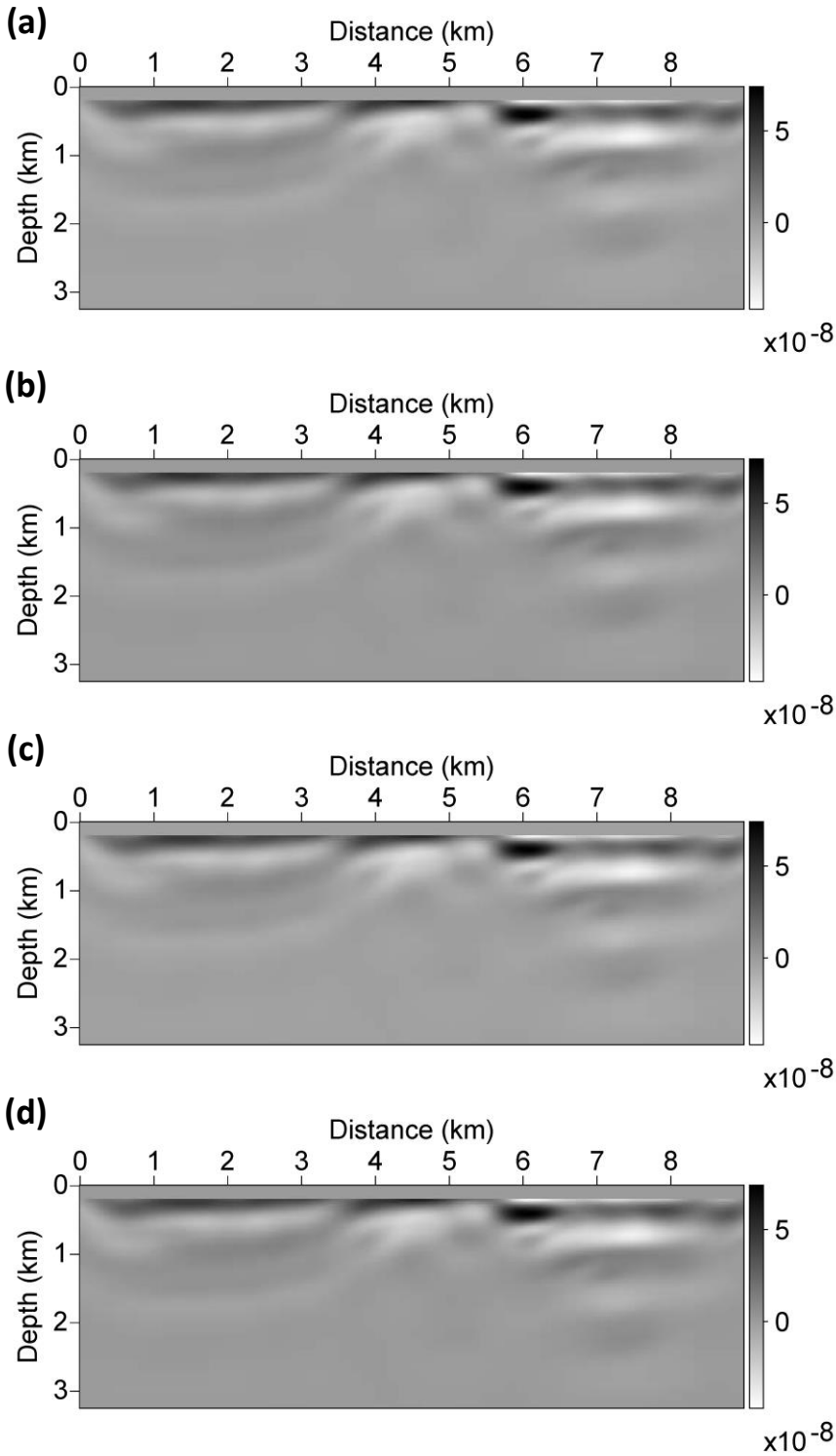

Figure 3.17 The matrix-vector product of the Hessian matrix and column vector $\mathbf{H}_i\mathbf{p}$ at 1.5 Hz obtained by (a) the improved HF method and by the conventional HF method with intervals of (b) 0.1, (c) 0.01 and (d) 0.001 for the Marmousi model.

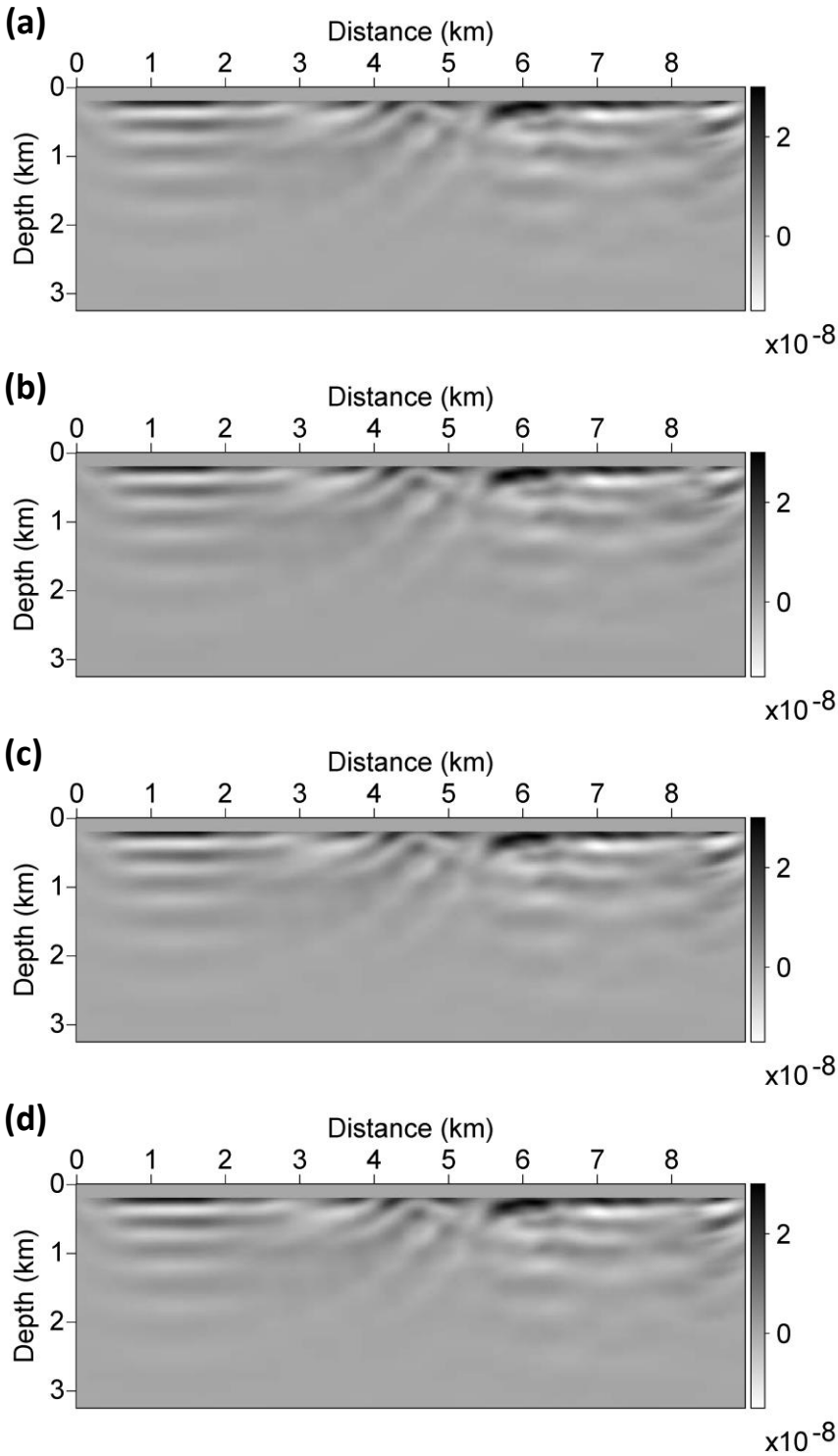

Figure 3.18 The matrix-vector product of the Hessian matrix and column vector $\mathbf{H}_i\mathbf{p}$ at 2.5 Hz obtained by (a) the improved HF method and by the conventional HF method with intervals of (b) 0.1, (c) 0.01 and (d) 0.001 for the Marmousi model.

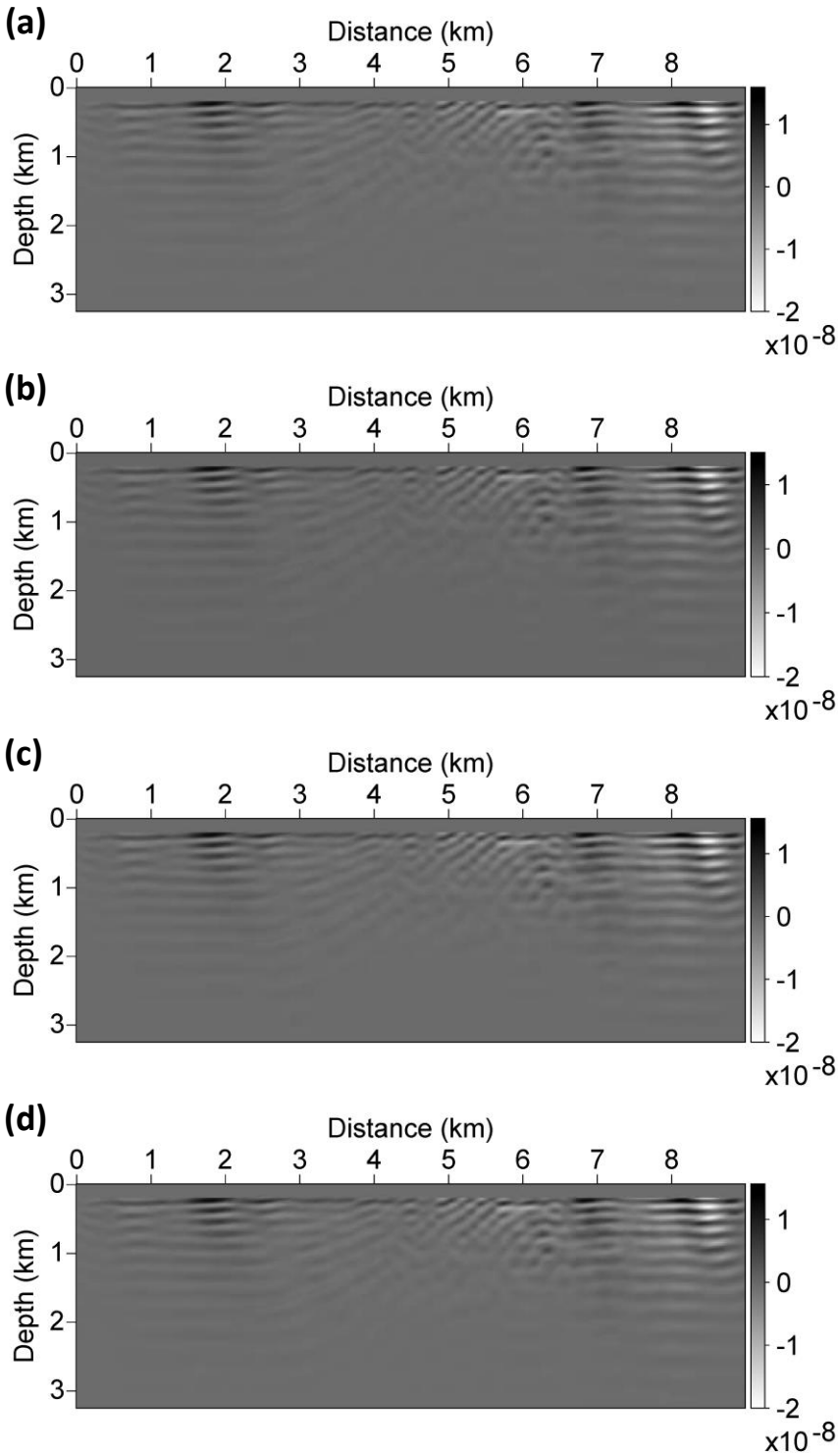

Figure 3.19 The matrix-vector product of the Hessian matrix and column vector $\mathbf{H}_i\mathbf{p}$ at 5.5 Hz obtained by (a) the improved HF method and by the conventional HF method with intervals of (b) 0.1, (c) 0.01 and (d) 0.001 for the Marmousi model.

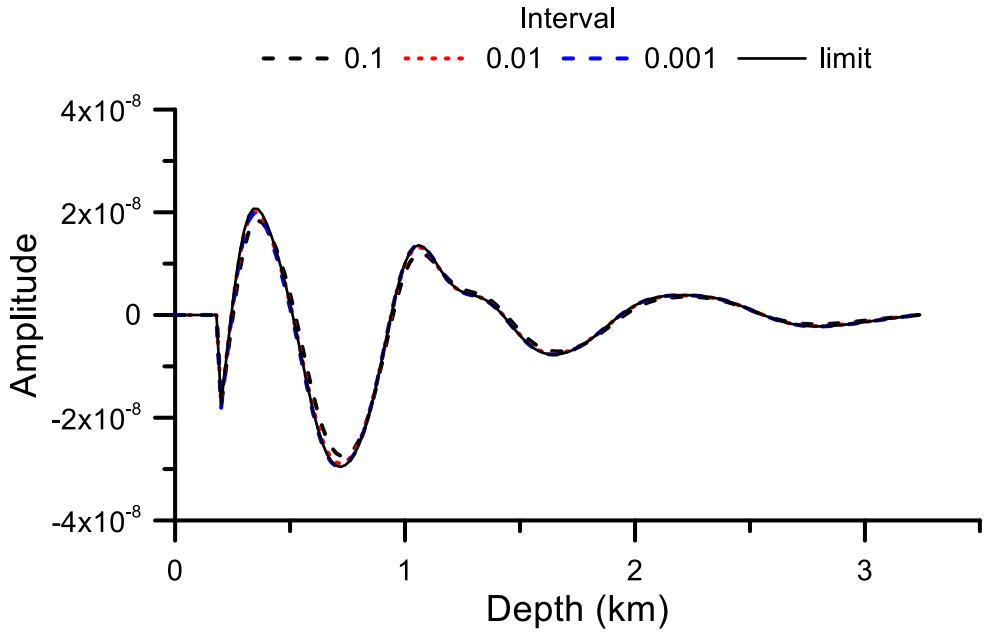

Figure 3.20 Depth profiles of $\mathbf{H}_i\mathbf{p}$ approximation obtained at a distance of 6.9 km for 1.5 Hz by the improved HF method using the limit of a function (black solid line) and by the conventional HF method using the forward FDA method with intervals of 0.1 (black dashed line), 0.01 (red dotted line) and 0.001 (blue dashed line).

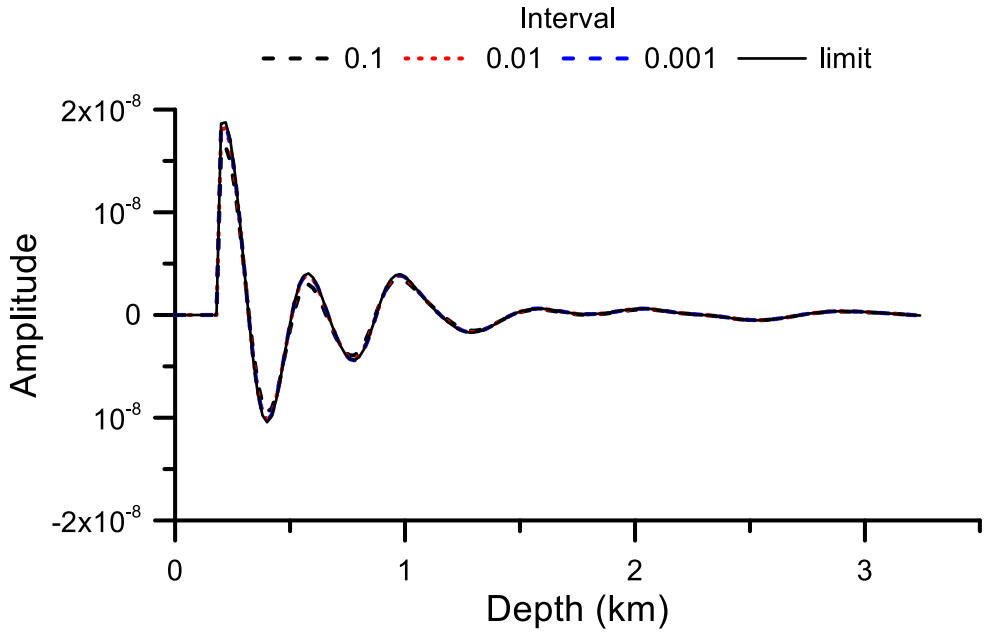

Figure 3.21 Depth profiles of $\mathbf{H}_i\mathbf{p}$ approximation obtained at a distance of 6.9 km for 2.5 Hz by the improved HF method using the limit of a function (black solid line) and by the conventional HF method using the forward FDA method with intervals of 0.1 (black dashed line), 0.01 (red dotted line) and 0.001 (blue dashed line).

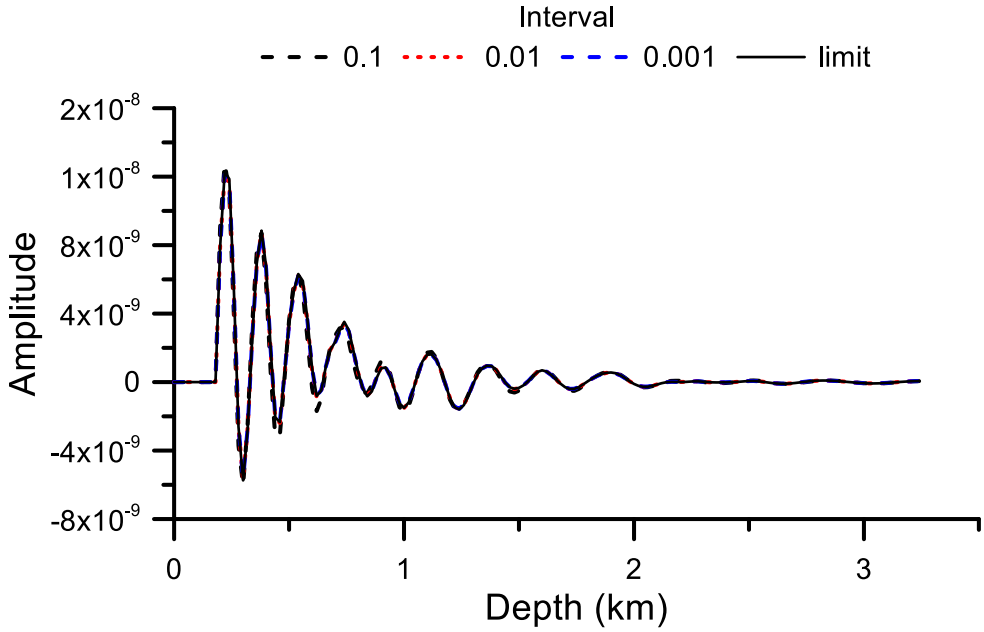

Figure 3.22 Depth profiles of $\mathbf{H}_i\mathbf{p}$ approximation obtained at a distance of 6.9 km for 5.5 Hz by the improved HF method using the limit of a function (black solid line) and by the conventional HF method using the forward FDA method with intervals of 0.1 (black dashed line), 0.01 (red dotted line) and 0.001 (blue dashed line).

Remember that the approximation intervals, which yield accurate and stable approximations of $\mathbf{Hp}$ for the Marmousi model, may not be proper for the acoustic Overthrust model (Figure 3.11). In other words, a number of numerical tests should be pre-performed before seismic FWI to determine an appropriate approximation interval depending on models. In addition, the interval is also affected by the initial and k^{th} inverted models as shown in Figures 3.13 and 3.14. To investigate robustness of the improved HF optimization method for model parameters, numerical tests for the acoustic Overthrust model are performed with the same inversion settings as in Section 3.1 and compared with those obtained by the conventional HF optimization method. Figure 3.23e shows search directions at the first FWI iteration obtained by the improved HF optimization method with the linearly-increasing initial model ranging from 3 km/s to 6.5 km/s (which is slightly overestimated). To investigate the convergence rates, RMS error curves are compared in Figure 3.24. Figure 3.24 shows that the conventional HF optimization method diverges, whereas the improved HF optimization method converges. The initial model is slightly changed with better estimation, whose velocity ranges from 3 km/s to 6 km/s. RMS error curves for the better initial model are shown in Figure 3.25. Figure 3.25 shows that the improved HF optimization method converges as fast as the conventional HF optimization method.

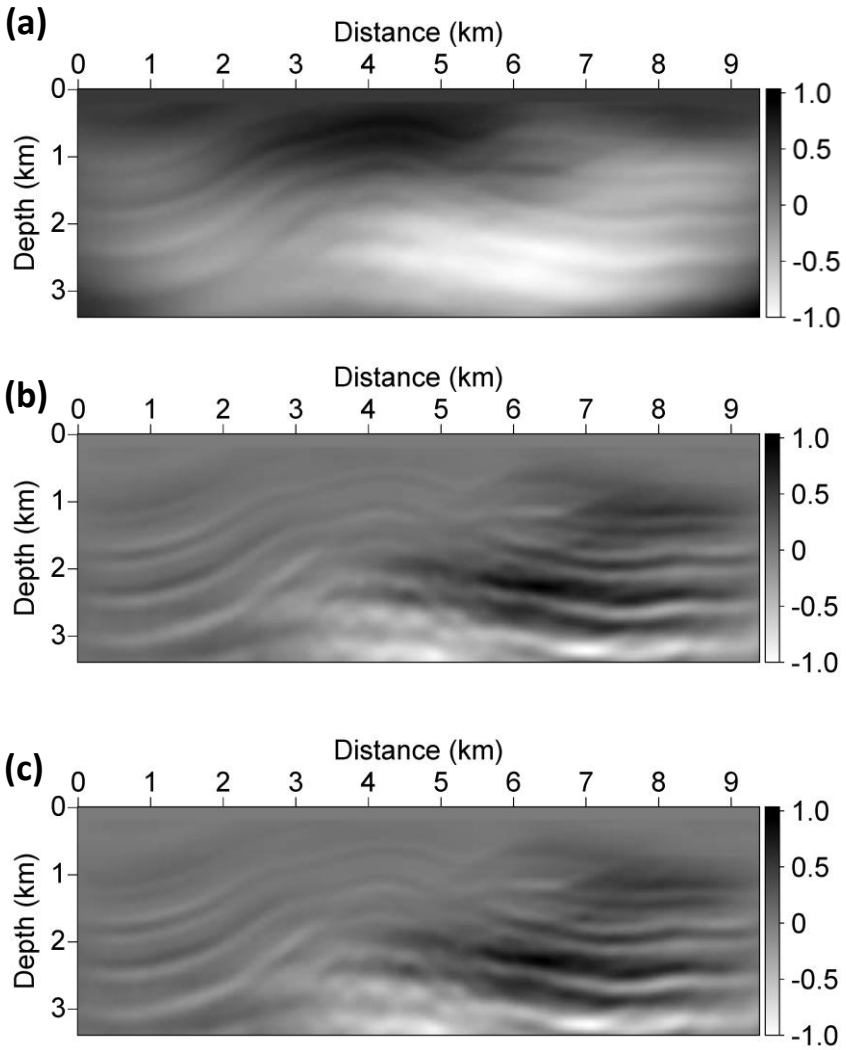

Figure 3.23 Comparison of search directions at the first FWI iteration obtained by (a) the improved HF method using the limit of a function and by the conventional HF method using the forward FDA method with intervals of (b) 0.1, (c) 0.01, (d) 0.001 and (e) 0.0001 for the acoustic Overthrust model.

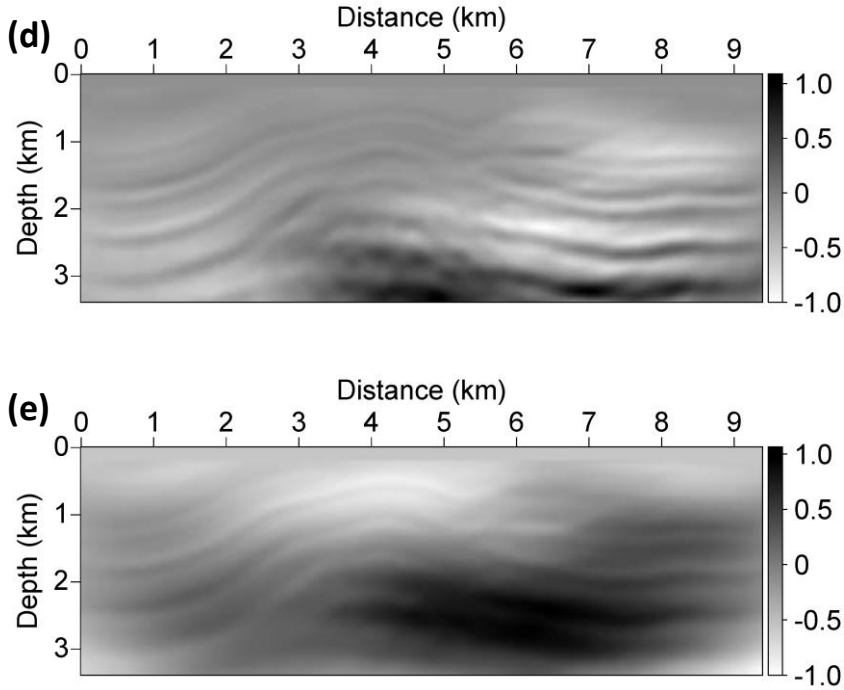

Figure 3.23 (Continued)

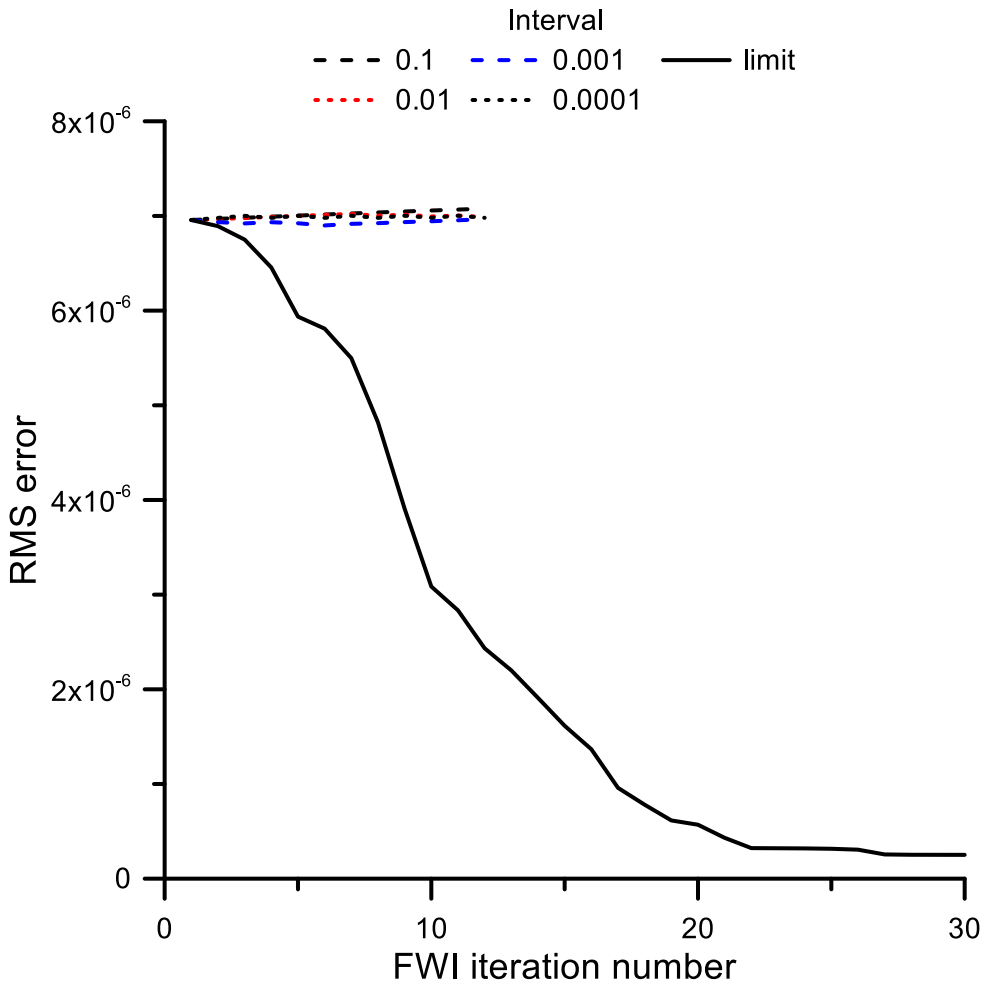

Figure 3.24 RMS error curves obtained by the improved HF method using the limit of a function (black solid line) and by the conventional HF method using the forward FDA method with intervals of 0.1 (black dashed line), 0.01 (red dotted line), 0.001 (blue dashed line) and 0.0001 (black dotted line).

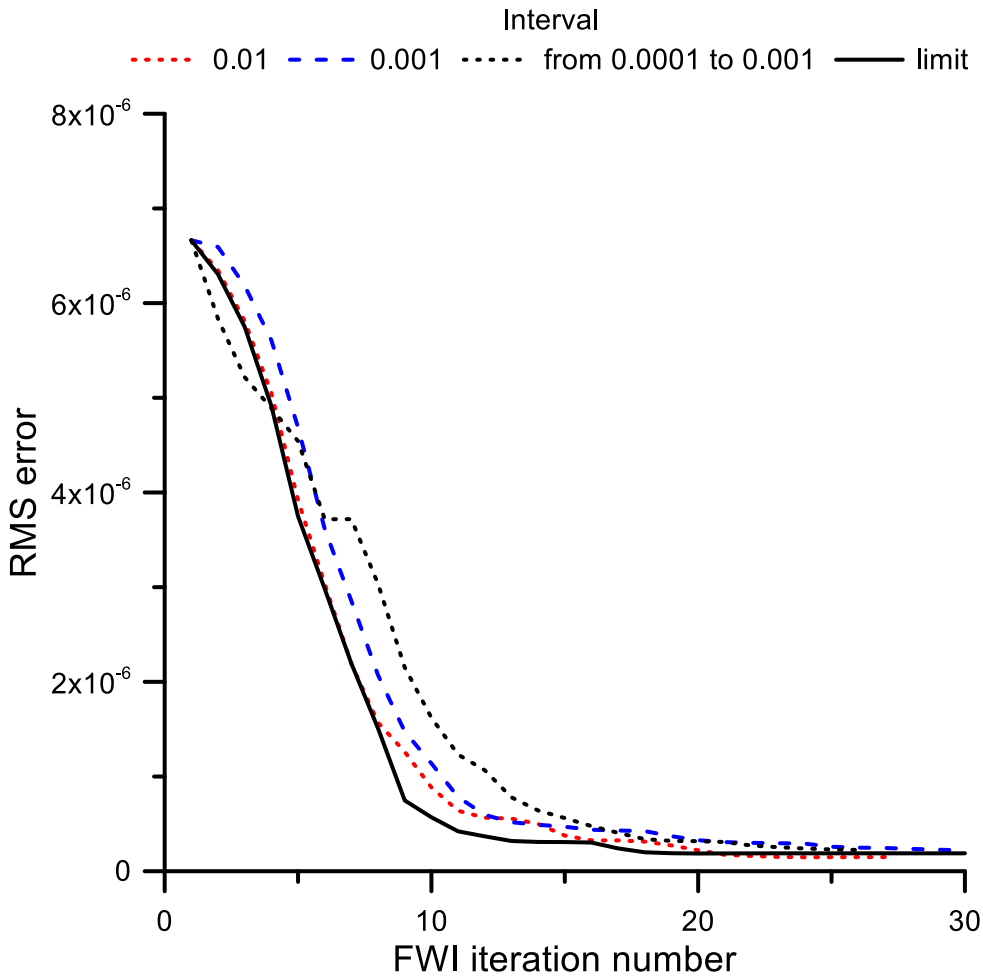

Figure 3.25 RMS error curves obtained by the improved HF method using the limit of a function (black solid line) and by the conventional HF method using the forward FDA method with intervals of 0.01 (red dotted line), 0.001 (blue dashed line) and intervals changing from 0.0001 to 0.001 (black dotted line).

3.2.3 Advantages of the improved Hessian-Free optimization

The improved HF optimization method was designed to approximate the matrix-vector product of the Hessian matrix and column vector in the linear conjugate-gradient loop. Its derivation was based not on the FDA method but on the limit of a function. In other words, a number of numerical tests for determining an appropriate approximation interval ε are not needed, because the approximation intervals are cancelled and go to 0 by the limit. In section 3.1, results of $\mathbf{Hp}$ approximations, which are approximated by the forward or central FDA methods, show that the FDA methods can be inaccurate depending on the approximation interval ε. In contrast, the improved HF optimization does not suffer from instability due to the approximation interval.

Another advantage of the improved HF optimization method is associated with computational cost. As mentioned in section 2.3.3, the conventional HF optimization using the forward FDA method carries out forward modeling twice per iteration for the linear conjugate-gradient method, whereas the central FDA method needs calculations of forward modeling 4 times. That's why the forward FDA method has been preferred to the central FDA method in seismic FWI fields, although the central FDA method yields more accurate and stable approximation. The improved HF optimization also performs forward modeling twice per iteration for the linear conjugate-gradient method just like the conventional HF optimization using the forward FDA method, one for calculating equation (3.26) and the other for equation (3.27). The main difference between the improved HF optimization and the conventional one using the central FDA is that the conventional HF method additionally construct and factorize the modeling operator for the perturbed modeled parameters [$\mathbf{S}\left(\mathbf{m}_0 + \varepsilon\mathbf{p}\right)$], whereas the improved HF optimization only applies the modeling operator $\mathbf{S}\left(\mathbf{m}_0\right)$, which was already constructed, factorized

and solved for the calculation of $\nabla E(\mathbf{m}_0)$. Because this process should be repeated at every iteration for the linear conjugate-gradient method, because the perturbed model parameter vector $\mathbf{p}$ varies over iterations. Accordingly, the improved HF method can achieve computational efficiency compared to the conventional HF method.

Chapter 4. Numerical examples

To demonstrate the stability and accuracy of the improved HF optimization method, the conventional and improved HF optimization methods are applied for the Marmousi and acoustic overthrust models. To guarantee the convergence of FWI, the approximation intervals for the conventional HF method are chosen through a number of pre-performed numerical tests. To accelerate the convergence rates and to improve inversion results, the preconditioning, weighting, source encoding and line-search techniques, which were already discussed in Chapter 2, are applied to both the optimization methods.

4.1. The Marmousi model

To compare inversion results obtained by the conventional and improved HF optimization methods for the Marmousi model, the approximation interval for the conventional HF optimization method is determined as 0.1, 0.01 and 0.001 by numerical tests which were already performed in Section 3.2.2. To avoid the multi-parameter problem, density is fixed at 1 g/cc and not inverted. Figure 4.1 shows P-wave velocity for the true Marmousi model and the linearly-increasing initial model ranging from 1.5 km/s to 4.5 km/s. Inversion settings for numerical examples are identical with Table 3.1. To guarantee convergence to the global minimum, the line-search technique is applied to determine the optimal step length of the search directions.

In Figure 4.2, RMS error curves obtained by the conventional and improved HF optimization methods are displayed. RMS error decreases at almost the same rate for both the conventional and improved HF optimization methods. This indicates that the improved HF optimization can be applied to seismic FWI just like the conventional HF optimization method. Considering that the approximation intervals for the conventional HF optimization method were chosen through a number of pre-performed numerical tests, the improved method can be better than the conventional method. Figure 4.3 shows inversion results for P-wave velocity and Figures 4.4, 4.5 and 4.6 show depth profiles recorded at distances of 2.3 km, 4.6 km and 6.9 km, respectively. Figures 4.3 to 4.6 show that the improved HF optimization method yields similarly good inversion results to those obtained by the conventional HF optimization method.

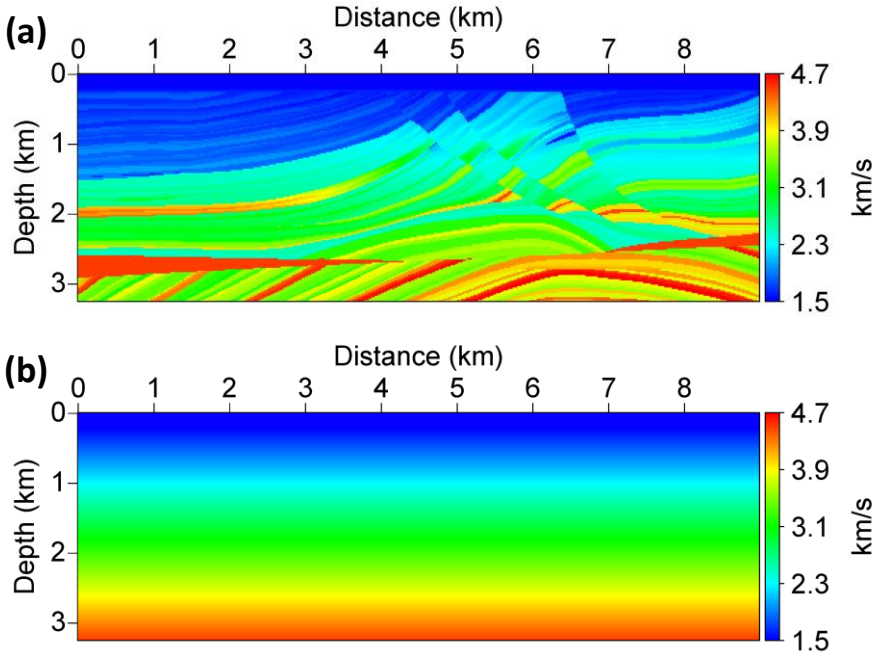

Figure 4.1 P-wave velocities of (a) the true Marmousi and (b) linearly-increasing initial models.

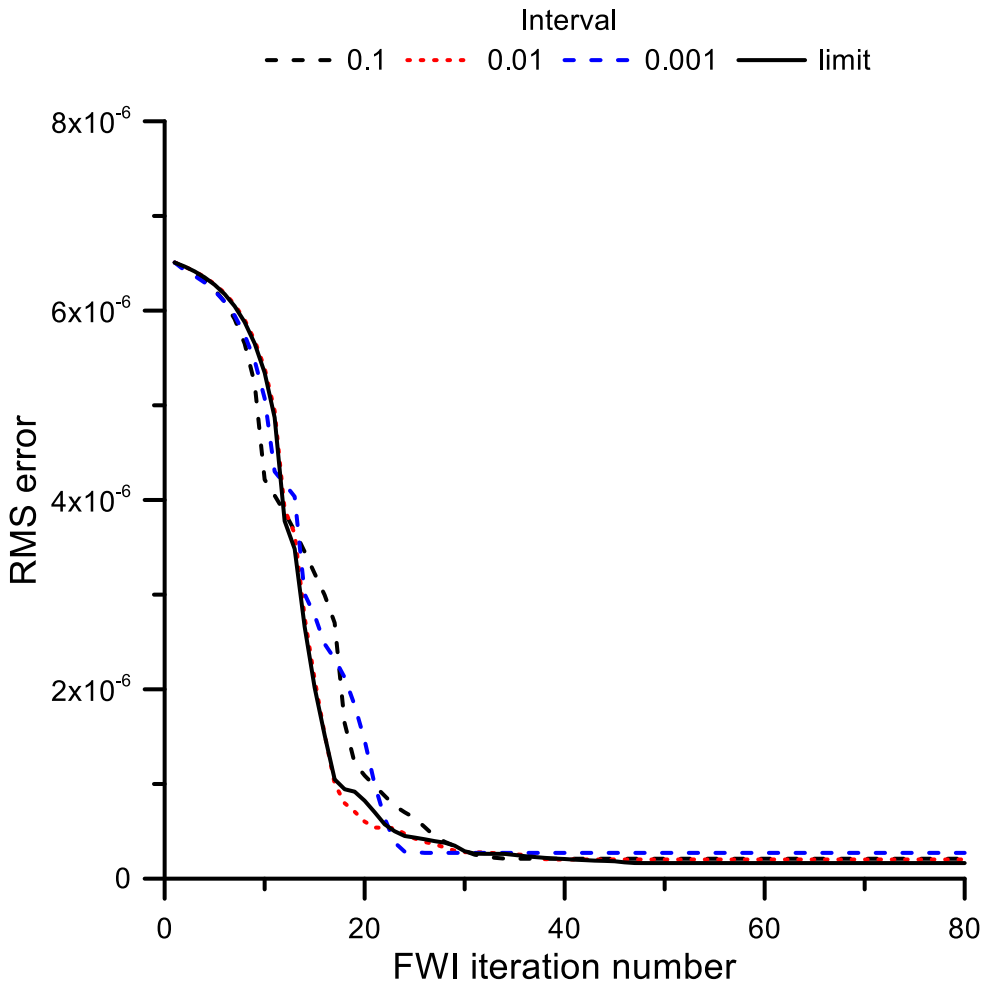

Figure 4.2 RMS error curves obtained by the conventional HF method using the forward FDA method with intervals of 0.1 (black dashed line), 0.01 (red dotted line), 0.001 (blue dashed line) and by the improved HF method using the limit of a function (black solid line).

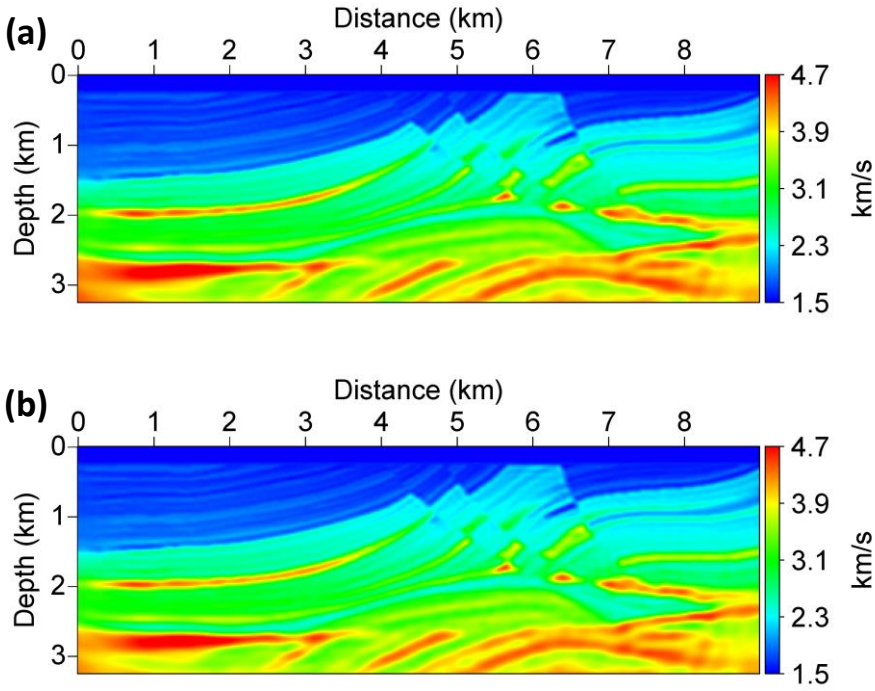

Figure 4.3 Seismic FWI results for P-wave velocity obtained by the conventional HF method using the forward FDA method with intervals of (a) 0.1, (b) 0.01 and (c) 0.001, and by (d) the improved HF method using the limit of a function.

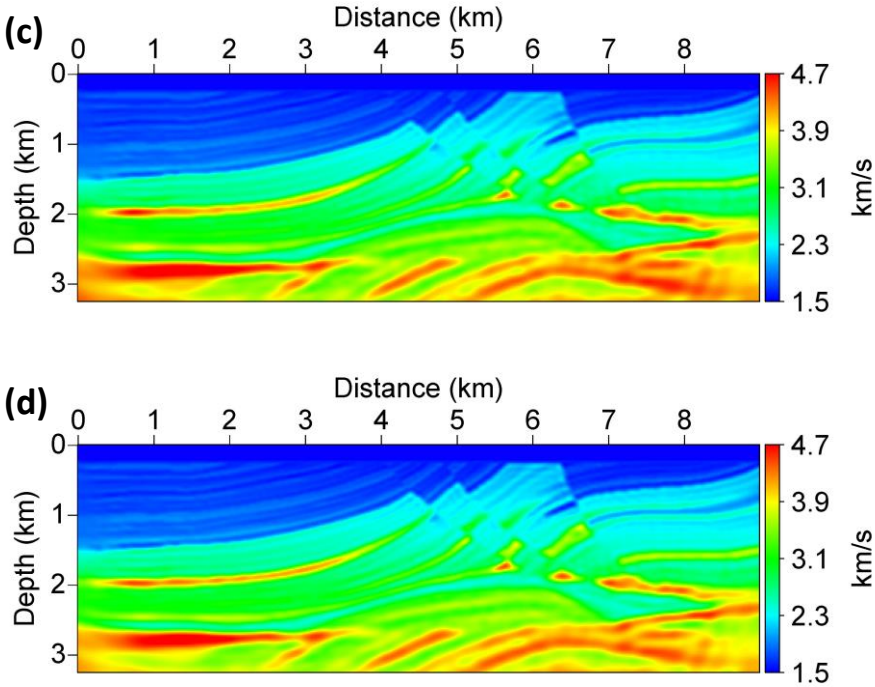

Figure 4.3 (Continued)

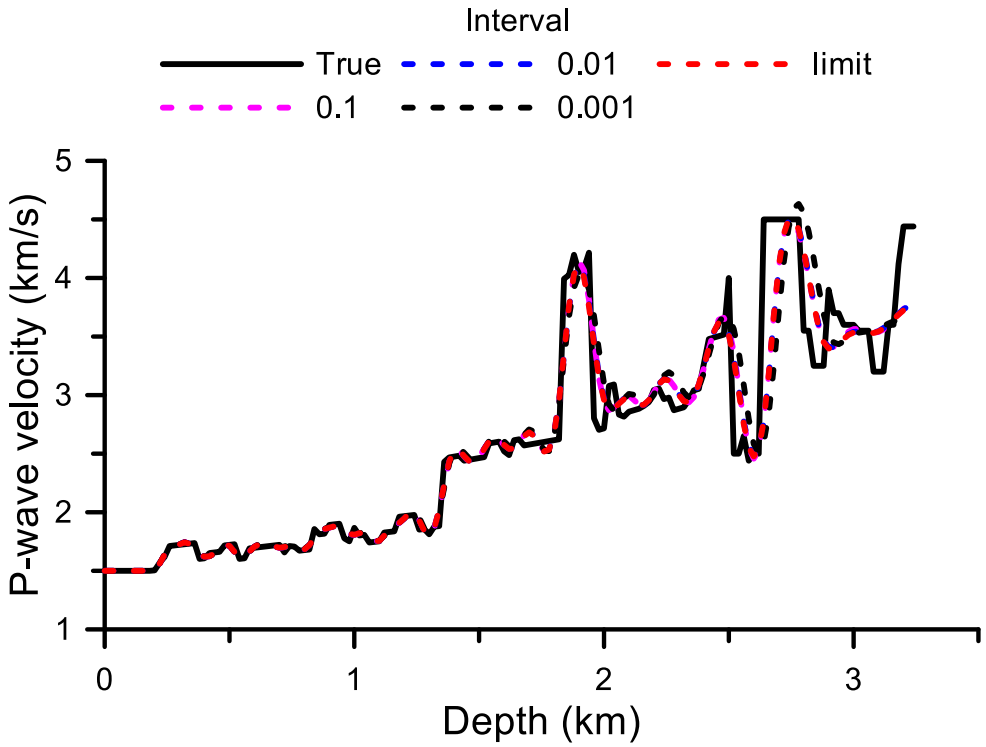

Figure 4.4 Depth profiles for P-wave velocity recorded at a distance of 2.3 km obtained by the conventional HF method using the forward FDA method with intervals of 0.1 (purple dashed line), 0.01 (blue dashed line) and 0.001 (black dashed line), and by the improved HF method using the limit of a function (red dashed line). The true velocities are denoted by the black solid line.

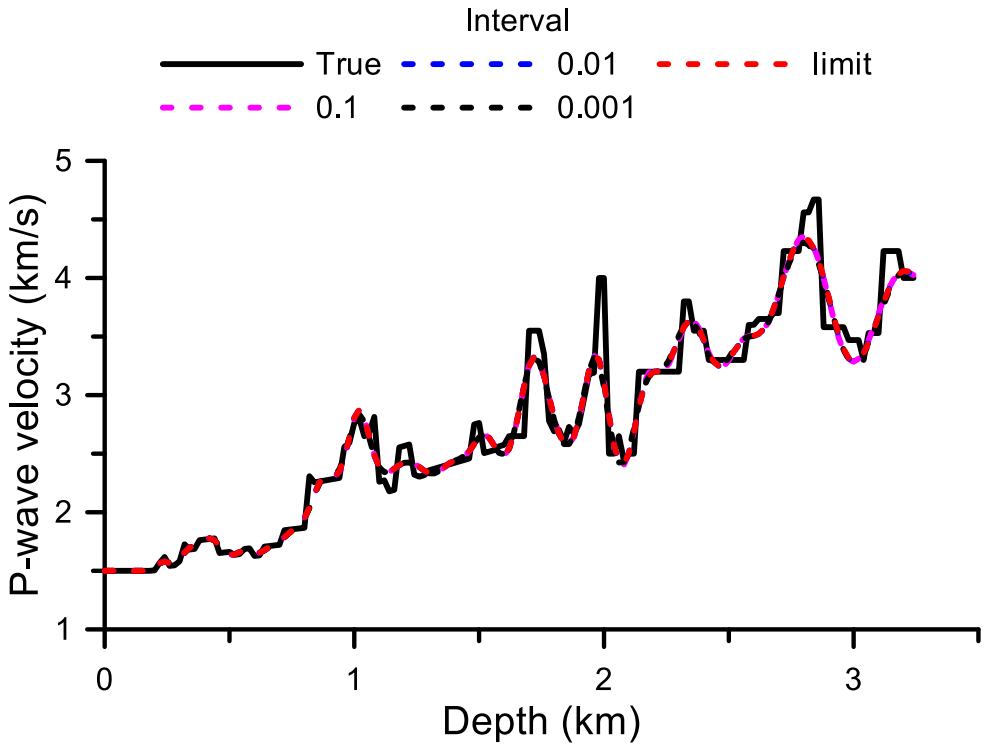

Figure 4.5 Depth profiles for P-wave velocity recorded at a distance of 4.6 km obtained by the conventional HF method using the forward FDA method with intervals of 0.1 (purple dashed line), 0.01 (blue dashed line) and 0.001 (black dashed line), and by the improved HF method using the limit of a function (red dashed line). The true velocities are denoted by the black solid line.

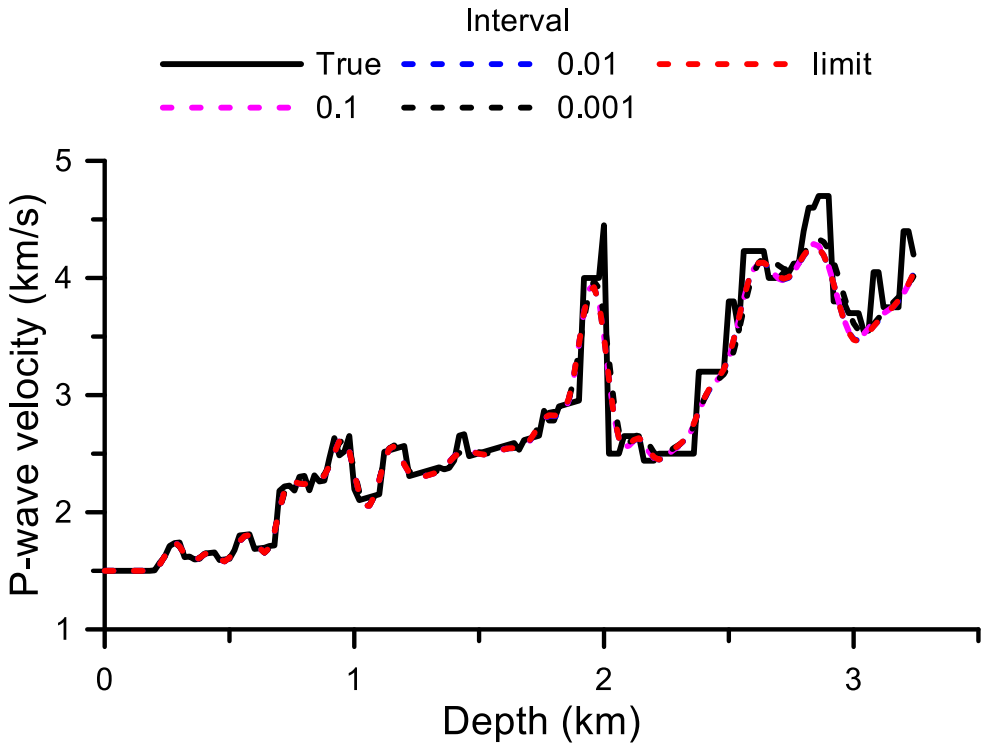

Figure 4.6 Depth profiles for P-wave velocity recorded at a distance of 6.9 km obtained by the conventional HF method using the forward FDA method with intervals of 0.1 (purple dashed line), 0.01 (blue dashed line) and 0.001 (black dashed line), and by the improved HF method using the limit of a function (red dashed line). The true velocities are denoted by the black solid line.

4.2. The acoustic Overthrust model with individual source

Numerical examples for the Marmousi model showed that the improved HF optimization method is applicable and reliable just like the conventional HF optimization method. In this section, numerical examples for the acoustic Overthrust model are presented and it is assumed that there are no low-frequency components below 3 Hz in observed data. The low-frequency components in observed data are essential to invert the long-wavelength structure in the early stage of seismic FWI (Sirgue and Pratt, 2004).

To compare the improved HF optimization method with the conventional HF optimization method, the approximation intervals of 0.1, 0.01 and 0.001 were chosen through a number of numerical tests to guarantee the accurate and stable approximation of the Hessian matrix. Table 4.1 shows the inversion settings for numerical examples and the line-search technique is also applied to enhance the convergence rate towards the global minimum. Figure 4.7 shows the true and initial models for the acoustic Overthrust model. RMS error curves are plotted in Figure 4.8. RMS error curve obtained by the conventional HF optimization method with an approximation interval of 0.1 shows different tendency at the 20^{th} iteration compared to the others. It indicates that using the approximation interval of 0.1 degrades the accuracy of the Hessian approximation. In other words, the conventional HF optimization method with approximation intervals larger than 0.1 may fail to converge to the global minimum. The failure of convergence in the seismic FWI process has been already discussed in Figures 3.24 and 3.25. Figure 4.9 shows inversion results for P-wave velocity obtained by the conventional and improved HF optimization methods. Figures 4.10, 4.11 and 4.12 show depth profiles for P-wave velocity recorded at distances of 2.4 km, 4.8 km and 7.2 km, respectively.

Table 4.1 Inversion settings for numerical examples of the acoustic Overthrust model.

Model size	No. of shot	No. of receiver	Shot interval	Recording time	Maximum Frequency	Minimum Frequency
9.6 km × **3.4 km**	157	470	0.06 km	8 sec	15 Hz	3 Hz

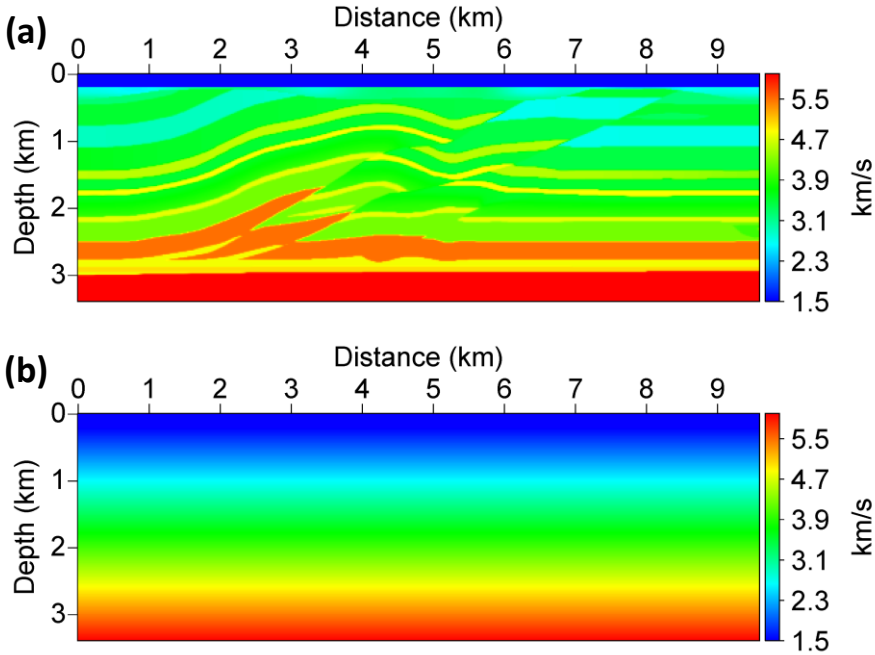

Figure 4.7 P-wave velocities of (a) the true acoustic Overthrust and (b) linearly-increasing initial models.

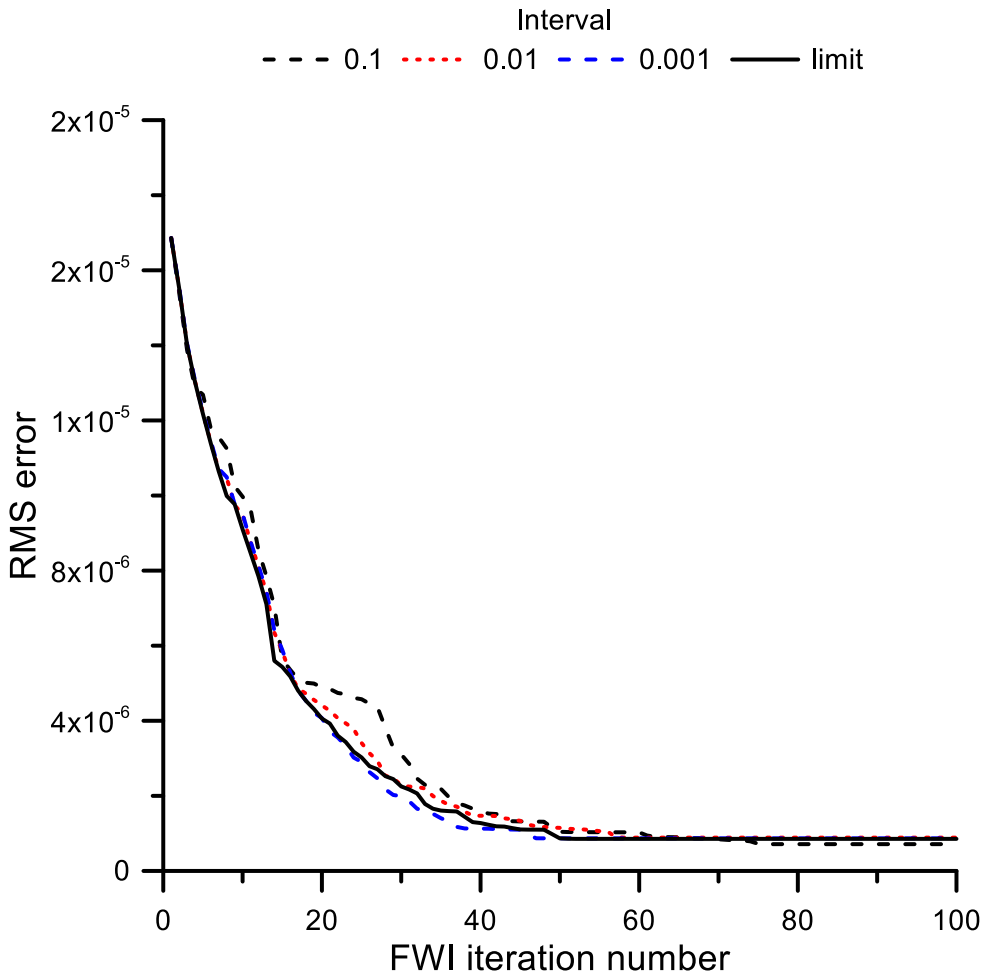

Figure 4.8 RMS error curves obtained by the conventional HF method with intervals of 0.1 (black dashed line), 0.01 (red dotted line) and 0.001 (blue dashed line), and by the improved HF method using the limit of a function (black solid line).

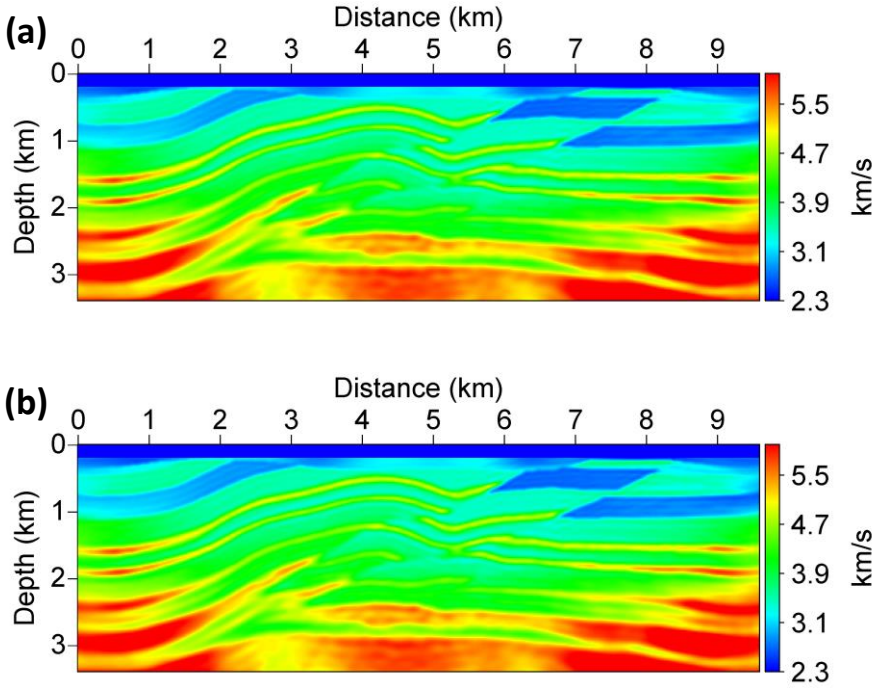

Figure 4.9 Seismic FWI results for P-wave velocity obtained by the conventional HF method using intervals of (a) 0.1, (b) 0.01 and (c) 0.001, and (d) by the improved HF method.

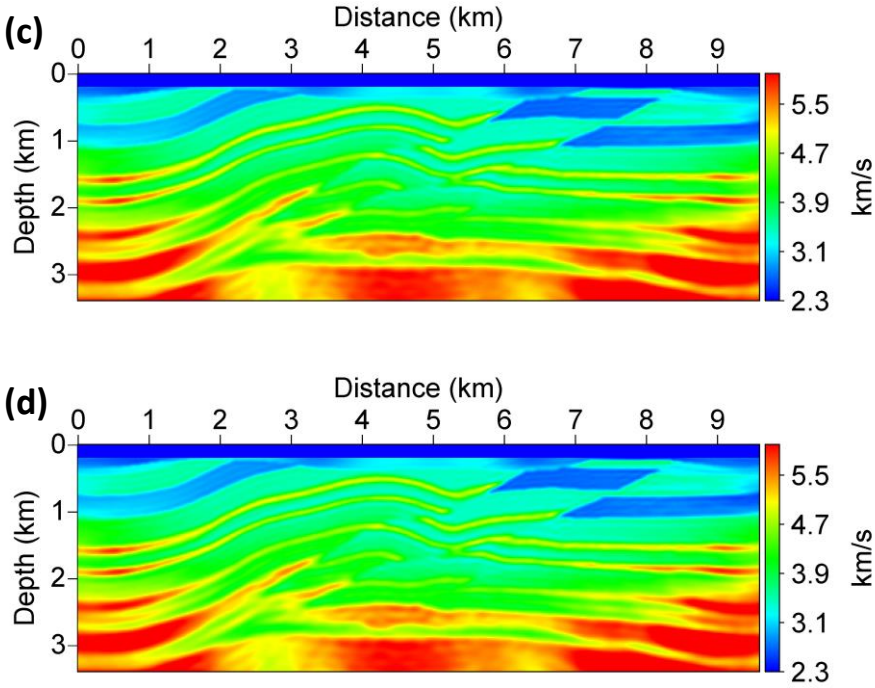

Figure 4.9 (Continued)

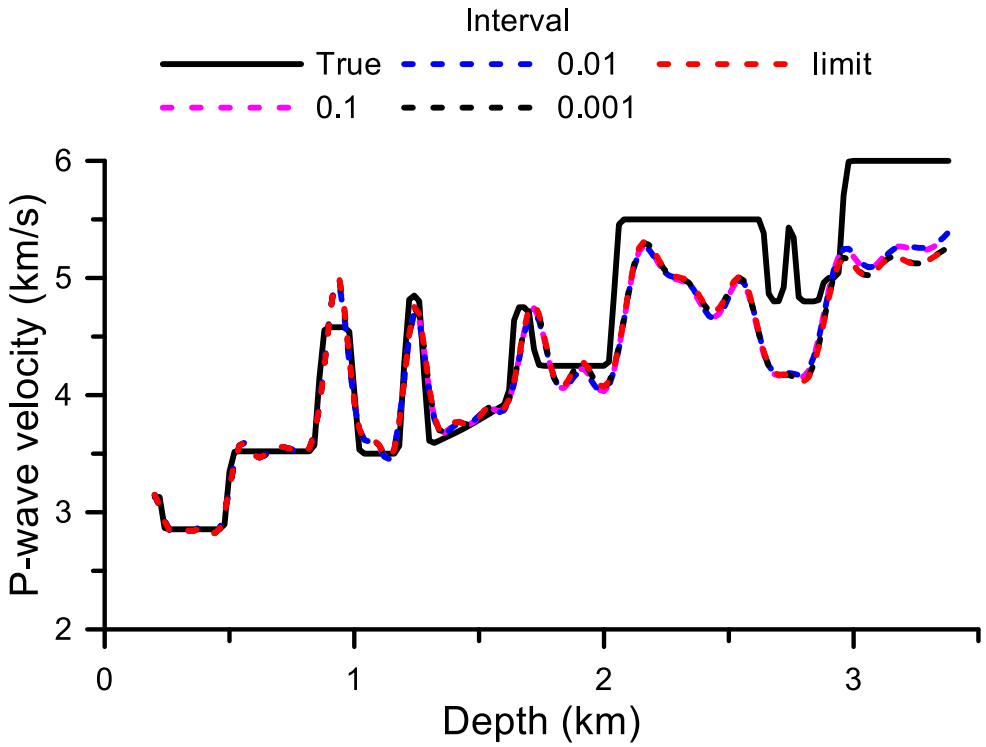

Figure 4.10 Depth profiles for P-wave velocity recorded at a distance of 2.4 km obtained by the conventional HF method using the forward FDA method with intervals of 0.1 (purple dashed line), 0.01 (blue dashed line) and 0.001 (black dashed line), and by the improved HF method using the limit of a function (red dashed line). The true velocities are denoted by the black solid line.

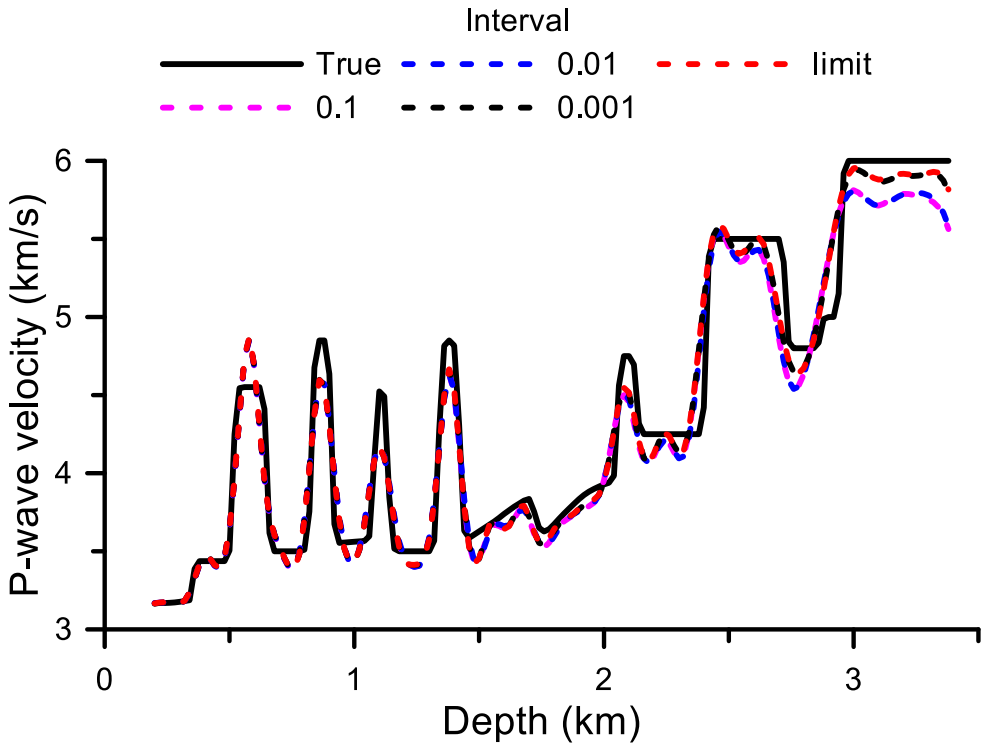

Figure 4.11 Depth profiles for P-wave velocity recorded at a distance of 4.8 km obtained by the conventional HF method using the forward FDA method with intervals of 0.1 (purple dashed line), 0.01 (blue dashed line) and 0.001 (black dashed line), and by the improved HF method using the limit of a function (red dashed line). The true velocities are denoted by the black solid line.

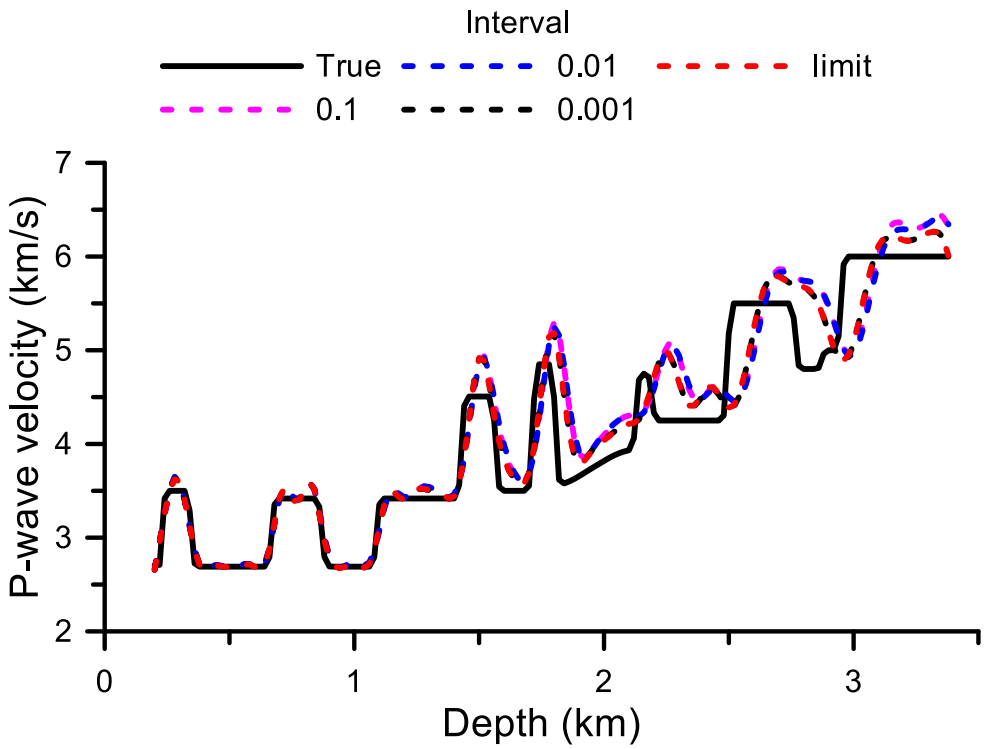

Figure 4.12 Depth profiles for P-wave velocity recorded at a distance of 7.2 km obtained by the conventional HF method using the forward FDA method with intervals of 0.1 (purple dashed line), 0.01 (blue dashed line) and 0.001 (black dashed line), and by the improved HF method using the limit of a function (red dashed line). The true velocities are denoted by the black solid line.

4.3. The acoustic Overthrust model with simultaneous source

The HF optimization method requires a number of forward modeling calculations to approximate the product of the Hessian matrix and column vector. In addition, computational cost of the forward modeling calculation depends on the number of seismic sources. To enhance computational efficiency, the source-encoding technique for the HF optimization method, which is proposed by Castellanos et al. (2015), is applied to the acoustic Overthrust model. The inversion settings are the same as those in Section 4.2.

As mentioned in Section 3.2.3, the improved HF optimization method does not need to additionally construct, factorize and solve the modeling operator $\mathbf{S}\left(\mathbf{m}_0+\varepsilon\mathbf{p}\right)$ for the perturbed model at every iteration for the linear conjugate-gradient method. In other words, the conventional HF optimization method requires to construct, factorize and solve the modeling operator $\mathbf{S}\left(\mathbf{m}_0+\varepsilon\mathbf{p}\right)$ at every iteration for the linear-conjugate gradient method. Figures 4.13, 4.14 and 4.15 show computing time required by the conventional HF optimization method and by the improved HF optimization method depending on the number of the source-encoding group. Table 4.2 shows average computing time per FWI iteration required by the conventional and improved HF optimization methods with 4, 8 and 16 encoding groups. The ratios of average computing time between the conventional and improved HF method are 3.19, 2.34 and 1.59 with 4, 8 and 16 encoding groups, respectively. To compare them under the same conditions, the iteration number of the linear-conjugate gradient method is fixed at 5. Computing time of the improved HF optimization method is much smaller than those of the conventional HF optimization method, because construction and factorization of the modeling operator for the perturbed model are not needed for the improved HF optimization method. From Figures 4.16 and 4.17, it is noted that as the number of

source-encoding groups increases, computing time increases. However, in the aspect of convergence rate, as the number of the source-encoding groups increases, the convergence rate is enhanced in Figures 4.18 and 4.19, since more and more source-encoding groups can suppress the crosstalk terms in the search direction of P-wave velocity as shown in Figure 4.20. There exists the trade-off between the number of source-encoding groups and convergence rate. Figures 4.21 and 4.22 show inversion results for P-wave velocity and RMS error curve obtained by the improved HF optimization method using the source-encoding and line-search techniques.

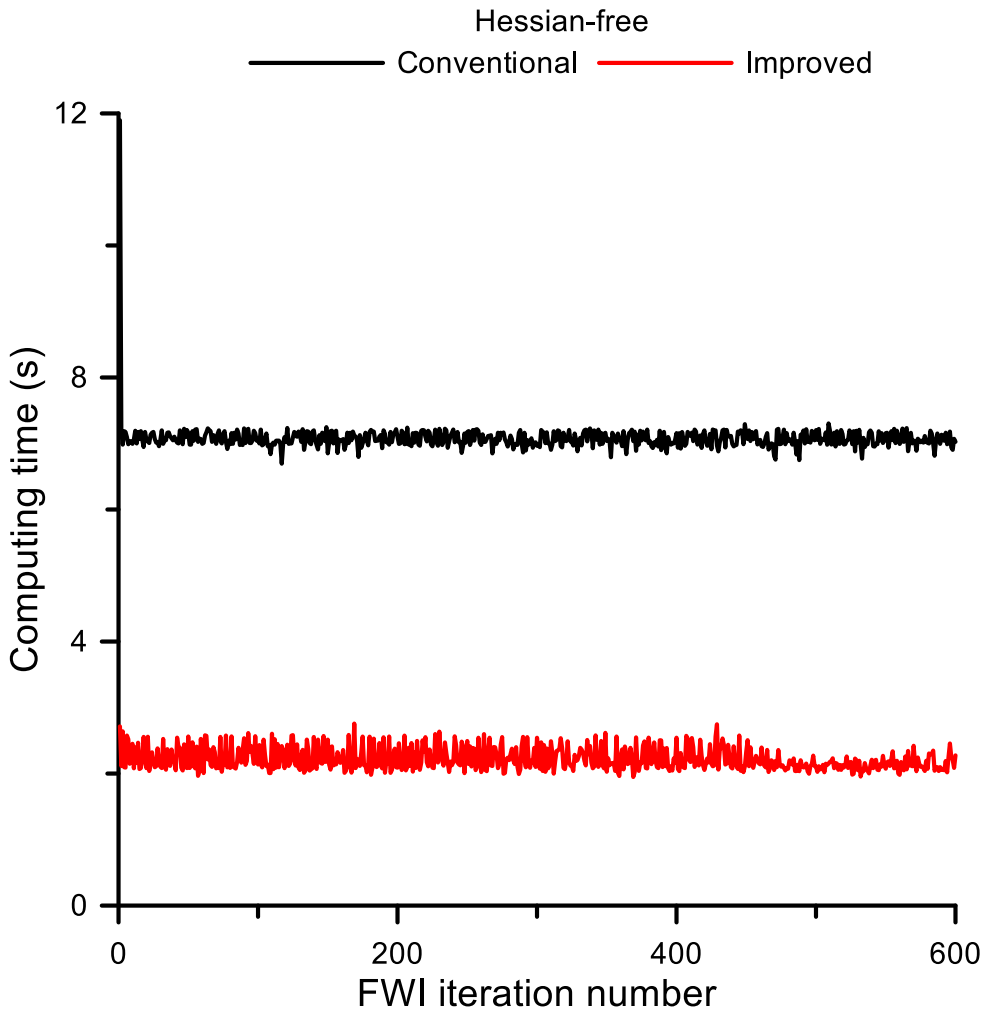

Figure 4.13 Computing time required by the conventional (black solid line) and improved (red solid line) HF optimization methods using the source-encoding technique with 4 encoding groups.

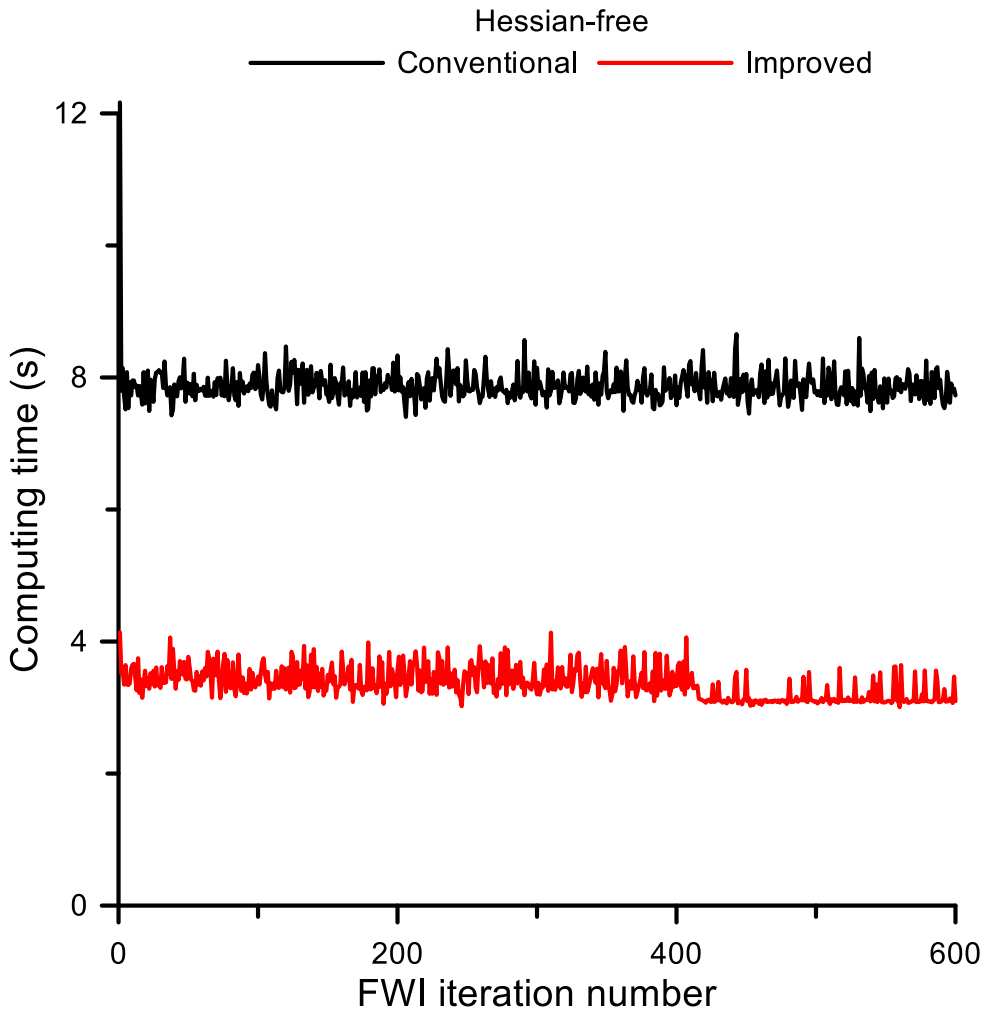

Figure 4.14 Computing time required by the conventional (black solid line) and improved (red solid line) HF optimization methods using the source-encoding technique with 8 encoding groups.

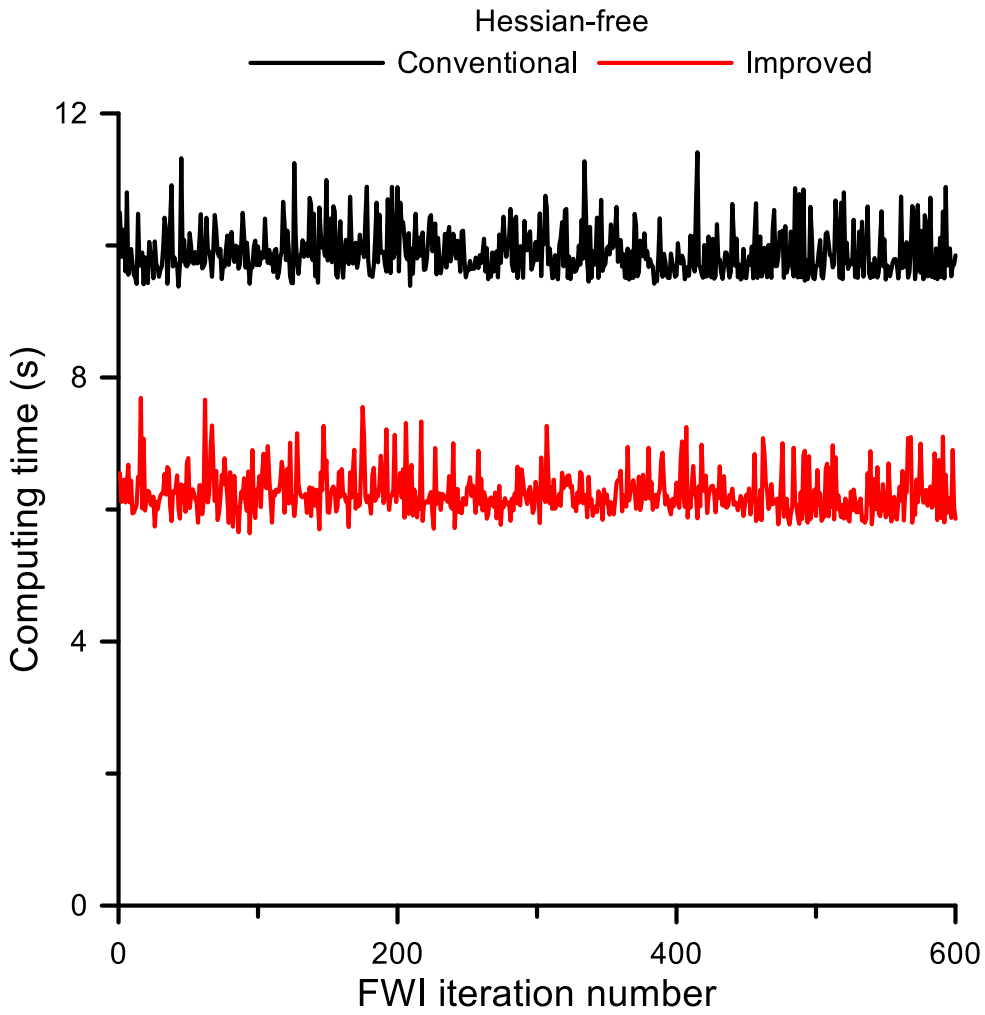

Figure 4.15 Computing time required by the conventional (black solid line) and improved (red solid line) HF optimization methods using the source-encoding technique with 16 encoding groups.

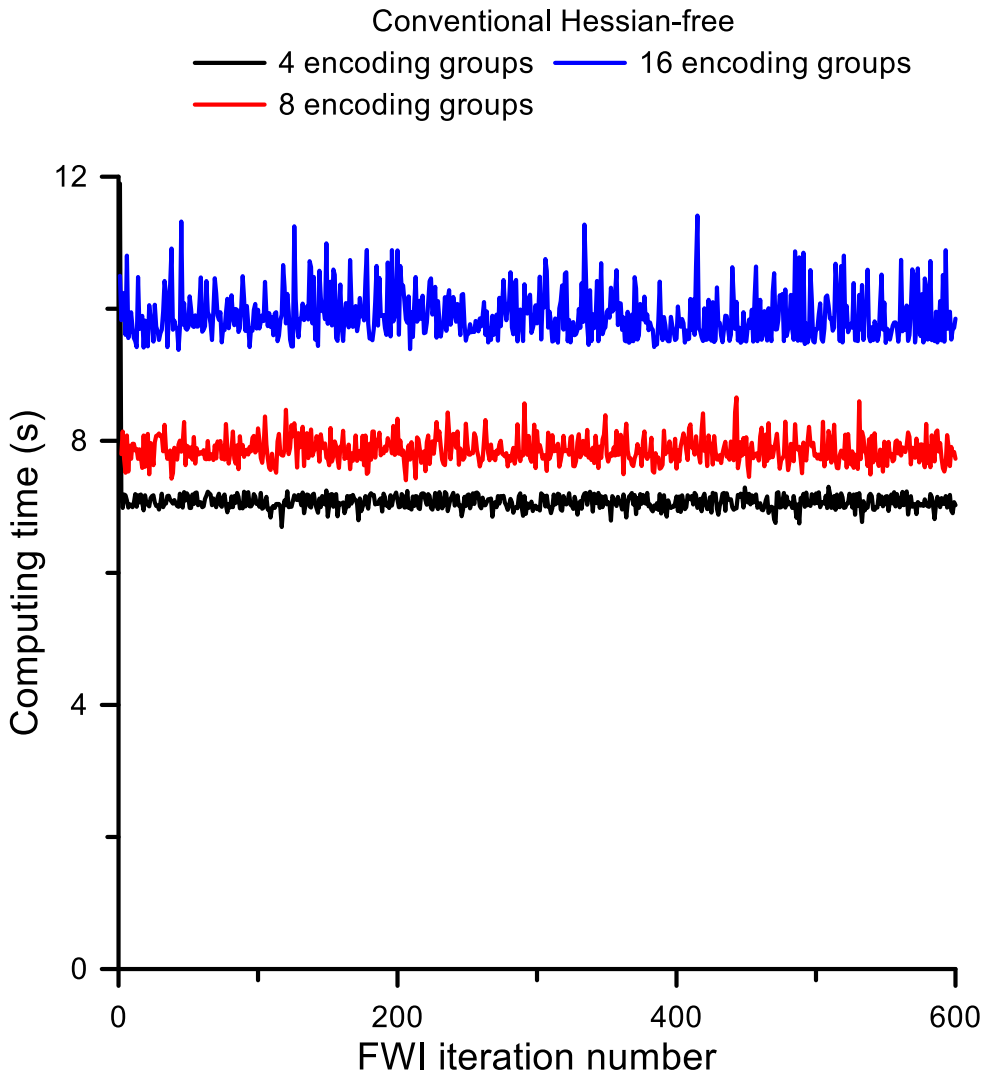

Figure 4.16 Computing time required by the conventional HF method using the source-encoding technique with 4 (black solid line), 8 (red solid line) and 16 (blue solid line) encoding groups.

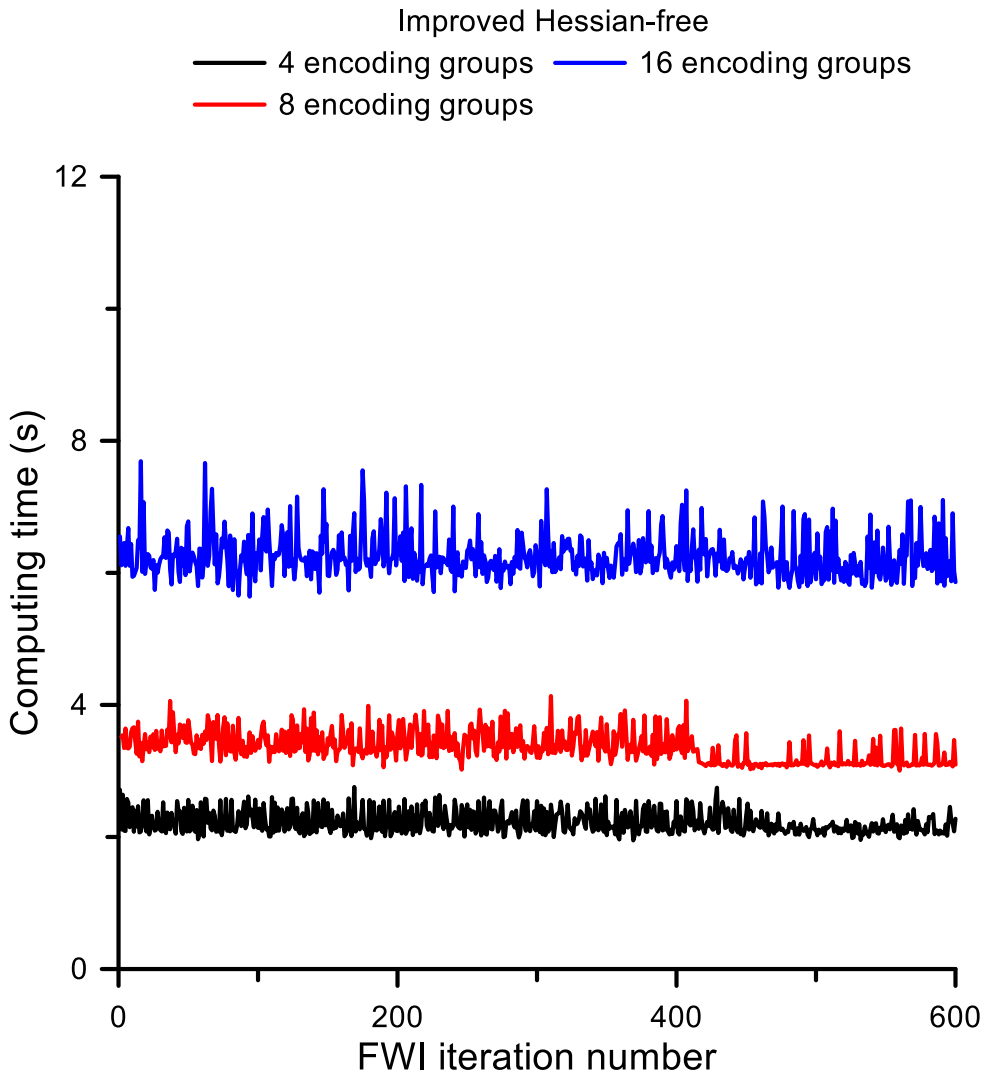

Figure 4.17 Computing time required by the improved HF method using the source-encoding technique with 4 (black solid line), 8 (red solid line) and 16 (blue solid line) encoding groups.

Table 4.2 Average computing time per FWI iteration required by the conventional and improved HF optimization methods using the source-encoding technique with 4, 8 and 16 encoding groups for the acoustic Overthrust model. Note that the ratio of the conventional HF method to the improved HF method decreases as the number of encoding group increases.

	4 encoding groups	8 encoding groups	16 encoding groups
Conventional HF	7.08 s	7.87 s	9.90 s
Improved HF	2.22 s	3.36 s	6.24 s
Ratio (C/I)	3.19	2.34	1.59

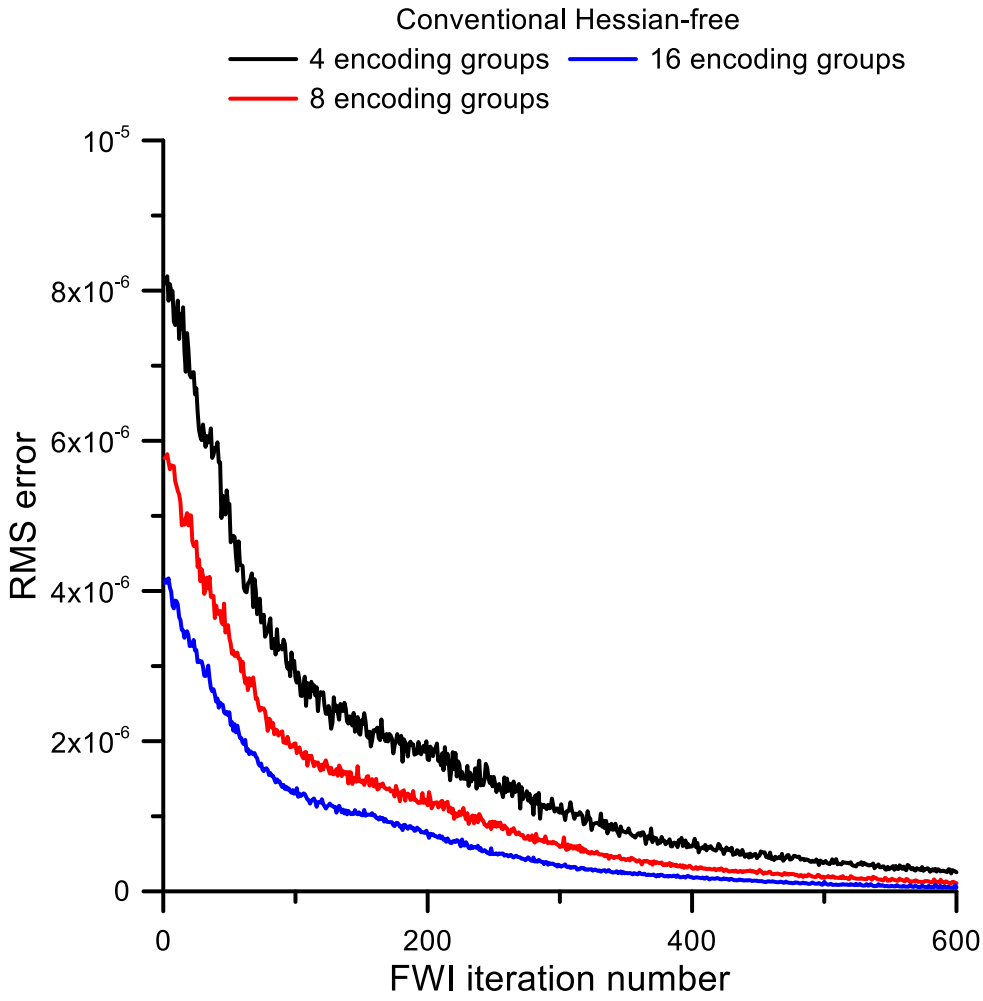

Figure 4.18 RMS error curves obtained by the conventional HF method using the source-encoding technique with 4 (black solid line), 8 (red solid line) and 16 (blue solid line) encoding groups. The approximation interval is 0.01.

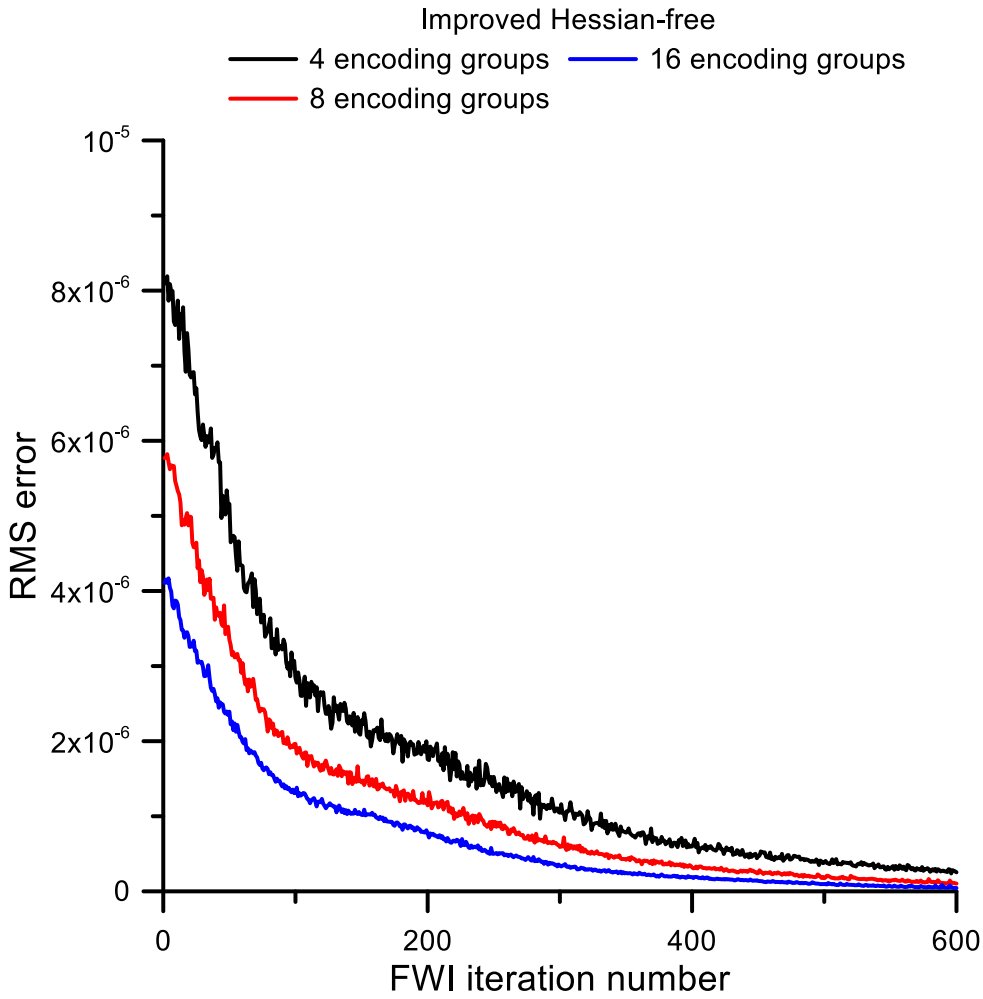

Figure 4.19 RMS error curves obtained by the improved HF method using the source-encoding technique with 4 (black solid line), 8 (red solid line) and 16 (blue solid line) encoding groups.

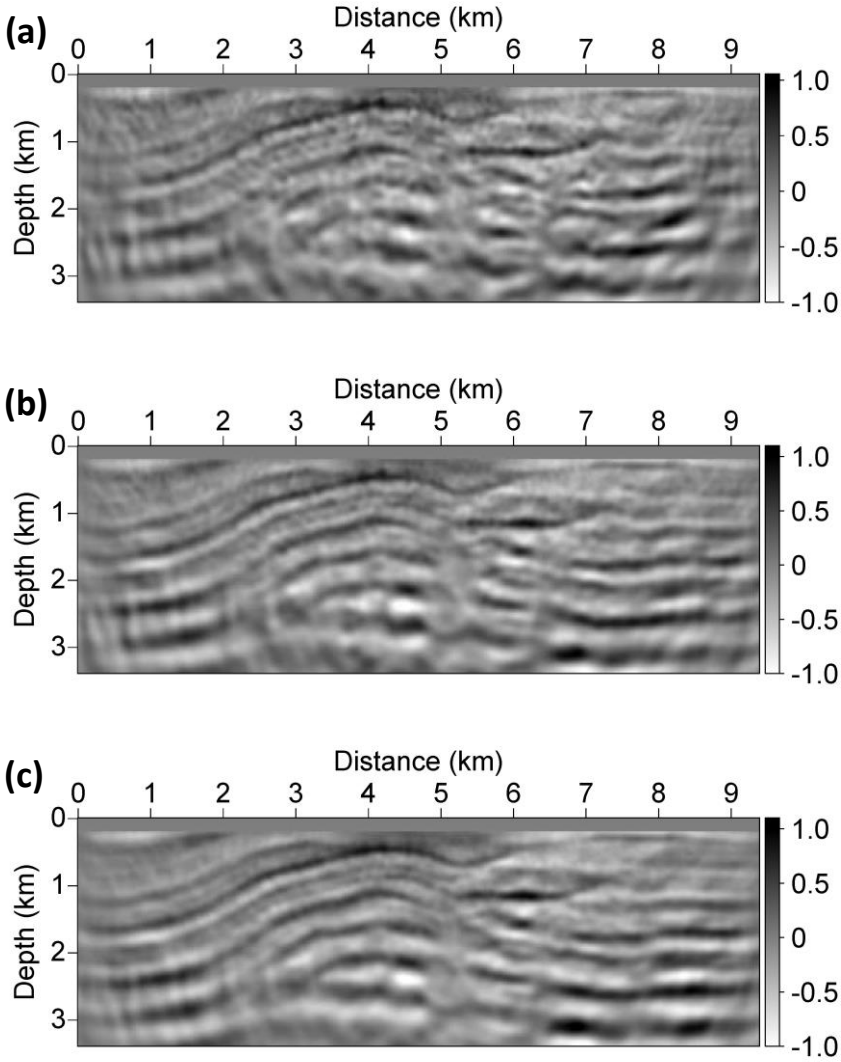

Figure 4.20 Search directions for P-wave velocity at the first FWI iteration obtained by the improved HF method using the source-encoding technique with source encoding group of (a) 4, (b) 8 and (c) 16.

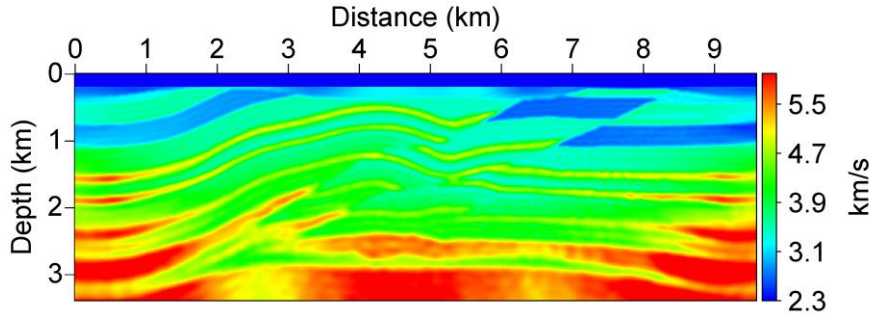

Figure 4.21 Seismic FWI results for P-wave velocity obtained by the improved HF method using the source-encoding technique with 4 groups and the line-search technique.

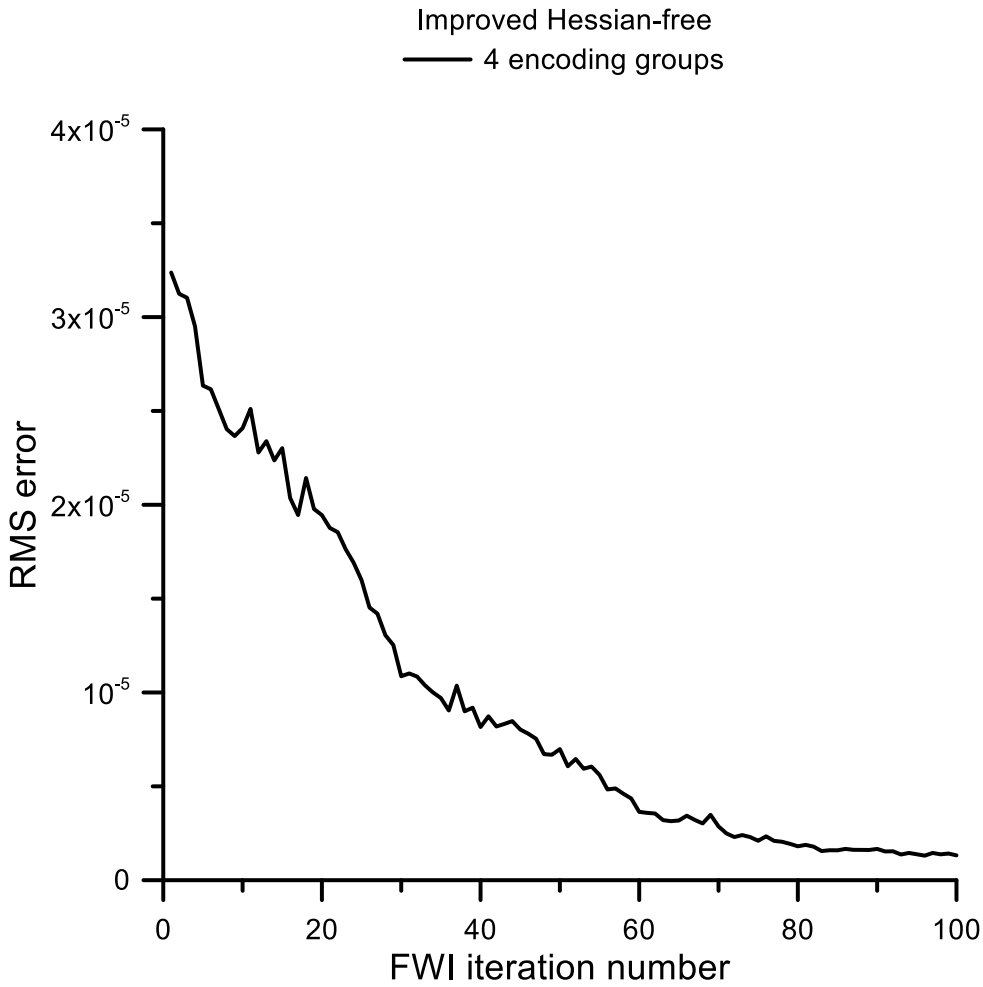

Figure 4.22 RMS error curve obtained by the improved HF method using the source-encoding technique with 4 groups and the line-search technique.

Chapter 5. Conclusions

Seismic full waveform inversion has been performed on the basis of the first-order (i.e., linear) optimization such as the steepest-descent method, because of its easy and intuitive implementation. However, the first-order optimization has a characteristic of focusing on recovering shallow structures in seismic FWI, which is a large-scale non-linear problem. To overcome this limitation, a preconditioner has been incorporated to the steepest-descent method, or the second-order optimization method using the Hessian matrix has been applied.

However, the Hessian matrix for seismic FWI requires too heavy computational cost and computing time with the present computer's capacities. To avoid calculating, storing and inverting the Hessian matrix explicitly, the Hessian-free optimization method is one of the most appealing and efficient optimization methods, because it does not require full information of the Hessian matrix itself. Instead of the Hessian matrix itself, the matrix-vector product of the Hessian matrix and column vector is needed in iterations for the linear conjugate-gradient method inside the seismic FWI code. The matrix-vector product of the Hessian matrix and column vector can be approximated by the second-order adjoint method or the FDA method. The FDA method has been popularly used to compute the matrix-vector product because it is simpler and more efficient to be implemented to seismic FWI than the second-order adjoint method. It only requires additional calculation of the steepest-descent directions for the perturbed model parameters to approximate the matrix-vector product of the Hessian matrix and column vector.

However, there exist several problems arising from the FDA method, which were discussed in Section 3.1. That is, the FDA method depends on the approximation interval ε. With too large or small approximation interval, the FDA method will fail in approximating the matrix-vector product of the Hessian

matrix and column vector due to the approximation errors $O(\varepsilon)$ or round-off errors, respectively. In addition, the appropriate approximation interval ε is dependent on model parameters, frequency components of data, and model perturbations. To choose an appropriate approximation interval ε, a number of the numerical tests are needed before seismic FWI. In addition, if one of the inversion environments changes during the inversion process, it would be better to find a new approximation interval ε.

To overcome those problems arising from the FDA method, the improved HF optimization was introduced in this study. The derivations of the new approximation method using the limit were addressed in Section 3.2. Its derivations are based on not the FDA method but the limit (i.e., the approximation interval ε approaches zero). Consequently, the improved HF optimization method approximates the matrix-vector product $\mathbf{Hp}$ without selection of an approximation interval ε. In other words, a number of the numerical tests are not needed to determine an appropriate approximation interval ε by using the new approximation method. In addition to the interval problem, the improved HF optimization method is more efficient than the conventional HF optimization method in the aspect of computational cost and computing time. Note that the conventional HF optimization method requires to construct, factorize and solve the new modeling operator $\mathbf{S}(\mathbf{m}_0 + \varepsilon\mathbf{p})$ for the perturbed model at every iteration for the linear conjugate-conjugate gradient method, whereas the improved HF optimization method only requires to solve the original modeling operator $\mathbf{S}(\mathbf{m}_0)$ which has been already constructed and factorized before the linear conjugate-gradient loop.

To demonstrate those advantages of the improved HF optimization method over the conventional HF method, numerical examples for the Marmousi and acoustic Overthrust models are provided and discussed in Chapter 4. Inversion results for P-wave velocity parameters obtained by the improved HF optimization

method show almost the same accuracy compared to those obtained by the conventional HF optimization method, although the approximation interval is not considered in the improved HF optimization method. In addition, the improved HF optimization method has proved to be much more efficient than the conventional HF method by comparing computing time when the source-encoding technique is incorporated.

The improved HF optimization can also be applied to other optimization problems in geophysics such as seismic FWI for elastic media and controlled-source electromagnetic (CSEM) inversion. Because those problems have similar inversion work flows such as construction of modeling operator and calculation of gradient schemes, the derivations of the improved HF method for elastic FWI and CSEM inversion will be similar to those for acoustic FWI. These applications will be addressed in further work.

Bibliography

Ben-Hadj-Ali, H., S. Operto, and J. Virieux, 2011, An efficient frequency-domain full waveform inversion method using simultaneous encoded sources, Geophysics, **76**, R109-R124.

Brossier, R., S. Operto, and J. Virieux, 2010, Which data residual norm for robust elastic frequency-domain full-waveform inversion?, Geophysics, **75**, R37-R46.

Castellanos, C., L. Metivier, S. Operto, R. Brossier, and J. Virieux, 2015, Fast full-waveform inversion with source encoding and second-order optimization methods: Geophysical Journal International, **200**, 718-744.

Choi, Y., Min, D. J. and Shin, C., 2008, Frequency-domain elastic full waveform inversion using the new pseudo-Hessian matrix: Experience of elastic Marmousi-2 synthetic data, Bulletin of the Seismological Society of America, **98**, 2402-2415.

Clayton, R. and B. Engquist, 1977, Absorbing boundary conditions for acoustic and elastic wave equations, Bulletin of the seismology society of America, **67**, 1529-1540.

Cohen, G., 2002, High-order numerical methods for transient wave equations, Springer.

Guitton, A., G. Ayeni, and E. Diaz, 2012, Constrained full-waveform inversion by model reparameterization, Geophysics, **77**, R117-R127.

Guitton, A., and Symes, W. W., 2003, Robust inversion of seismic data using the Huber norm, Geophysics, **68**, 1310-1319.

Ha, T., Chung, W. and Shin, C., 2009, Waveform inversion using a backpropagation algorithm and a Huber function norm, Geophysics, **74**, R15-R24.

Higdon, R. L., 1986, Absorbing boundary conditions for difference

approximations to the multidimensional wave equation, Mathematics of computation, **47**, 437-459.

Jang, U., Min, D.J. and Shin, C., 2009, Comparison of scaling methods for waveform inversion, Geophysical Prospecting, **57**, 49-59.

Jeong, W., S. Pyun, W. Son, and D.-J. Min, 2013, A numerical study of simultaneous source full waveform inversion with l_1-norm, Geophysical Journal International, **194**, 1727-1737.

Kelly, K. R., Ward, R. W., Treital, S., and Alford, R. M., 1976, Synthetic seismograms: a finite-difference approach, Geophysics, **41**, 2-27.

Krebs, J., J. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M. D. Lacasse, 2009, Fast full-wavefield seismic inversion using encoded sources, Geophysics, **74**, WCC177-WCC188.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations, Conference on inverse scattering, Theory and application: Society for Industrial and Applied Mathematics, Expanded abstracts, 206-220.

Levenberg, K., 1944, A method for the solution of certain non-linear problems in least squares, Quarterly of Applied Mathematics, **2**, 164-168.

Marfurt, K. J., 1984, Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations, Geophysics, **49**, 533-549.

Metivier, L., R. Brossier, J. Virieux, and S. Operto, 2013, Full waveform inversion and the truncated Newton method: SIAM Journal on Scientific Computing, **35**, B401-B437.

Metivier, L., R. Brossier, S. Operto, and J. Virieux, 2015, Acoustic multi-parameter FWI for the reconstruction of P-wave velocity, density and attenuation: preconditioned truncated Newton approach, 85th Annual International Meeting, SEG, Expanded Abstracts, 1198-1203.

Nash, S. G., 2000, A survey of truncated-Newton methods: Journal of

Computational and Applied Mathematics, **124**, 45-59.

Nocedal, J., and S. J. Wright, 1999, Numerical optimization, Springer, USA.

Nocedal, J., and S. J. Wright, 2006, Numerical optimization, Springer, USA.

Oh, J.W., and D.J. Min, 2013, Weighting technique using back-propagated wavefields incited by deconvolved residuals for frequency-domain elastic waveform inversion: Geophysical Journal International, **194**, 322-347.

Pan, W., Innanen, K. A. and Liao, W., 2017, Accelerating Hessian-free Gauss-Newton full-waveform inversion via l-BFGS preconditioned conjugated-gradient algorithm, Geophysics, **82**, R49-R64.

Plessix, R. –E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, Geophysical Journal International, **167**, 495-503.

Pratt, R. G., Shin, C., Hicks G. J., 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion, Geophysical Journal International, **133**, 341-362.

Pyun, S., Son, W., and Shin, C., 2009, Frequency-domain waveform inversion using an l_1-norm objective function, Exploration Geophysics, **40**, 227-232.

Reynolds, A., 1978, Boundary conditions for the numerical solution of wave propagation problems, Geophysics, **42**, 1099-1110.

Schiemenz, A., Lewis, W., Nunez, L. Z., Sabaa, A. E., Powell, S., Yu, M. and Imamshah, A., 2014, Improved imaging resolution using a hybrid l-BFGS – truncated Newton method in FWI: application to Bruce 3D field data, 84th Annual International Meeting, SEG, Expanded Abstracts, 971-975.

Schuster, G.T., X. Wang, Y. Huang, W. Dai, and C. Boonyasiriwat, 2011, Theory of multisource crosstalk reduction by phase-encoded statics, Geophysical Journal International, **184**, 1289-1303.

Shin, C., 1995, Sponge boundary condition for frequency-domain modeling, Geophysics, **60**, 1870-1874.

Shin, C., K. Yoon, K. J. Marfurt, K. Park, D. Yang, H. Y. Lim, S. Chung, and S.

Shin, 2001, Efficient calculation of a partial-derivative wavefield using reciprocity for seismic imaging and inversion, Geophysics, **66**, 1856-1863.

Shin, C., and Min, D.J., 2006, Waveform inversion using a logarithmic wavefield, Geophysics, **71**, R31-R42.

Sirgue, L., and Pratt, R. G., 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies, Geophysics, **69**, 231-248.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation, Geophysics, **49**, 1259-1266.

Virieux, J. and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics, Geophysics, **74**, WCC127-WCC152.

Wang, Y., Dong, L. and Wang, Z., 2016, FWI using truncated Newton method based on high performance computing, 2016 SEG High Performance Computing Workshop, 37-38.

Zienkiewicz, O. C., and Taylor R. L., 2000, The finite element method, volume 1: The basis Butterworth-Heinemann.

국문초록

음향파 파형역산을 위한
개선된 헤시안-프리 최적화 기법

탄성파 파형역산은 지하 매질의 물성 정보를 추출하는 탄성파 자료처리 기법 중 하나로, 관측 파동장과 수치 파동장의 잔차를 최소화하는 방향으로 물성 정보를 반복적으로 업데이트 한다. 잔차를 최소화하는 물성 방향을 구하기 위한 여러 최적화 기법 중, 본 연구에서는 헤시안-프리 최적화 기법을 이용한다. 헤시안-프리 최적화 기법은 헤시안 행렬을 직접 계산하거나 그 역행렬을 계산할 필요가 없기 때문에 다른 최적화 기법에 비해 계산량이 적으며 탄성파 파형역산과 같이 많은 물성 값을 갖는 비선형 문제에 적합한 기법이다. 직접적인 헤시안 행렬 계산 대신, 헤시안-프리 최적화 기법은 헤시안 행렬과 임의의 열벡터 곱이 켤레기울기법 과정에서 필요하며, 이는 유한차분법을 이용하면 쉽게 근사값을 구할 수 있다. 그러나 정확하고 안정적으로 근사값을 구하기 위해서는 적절한 차분간격을 반복적인 수치 예제를 통해 구해야 하며, 지하 매질의 물성이나 주파수와 같은 다양한 역산 변수들에 따라 새로운 차분간격을 다시 구해야 하는 단점이 있다.

본 연구에서는 이러한 문제를 해결하기 위해 유한차분법이 아닌 극한값을 이용하는 개선된 헤시안-프리 최적화 기법을 제안한다. 극한의 정의에 따라 새로운 방법은 차분간격이 0에 가까워지며 이로 인해 역산 환경에 따라 매번 새로운 차분간격을 구할 필요가 없이 정확하고 안정으로 헤시안 행렬과 열벡터의 곱을 구할 수 있다는 장점이 있다. 또한 개선된 헤시안-프리 최적화 기법은 기존의 헤시안-프리

최적화 기법과 달리 켤레기울기법 과정에서 새로운 임피던스 행렬을 구축하거나 분해할 필요가 없기 때문에 훨씬 빠르게 파형역산을 수행 할 수 있다. 따라서, 본 연구에서 제안하는 개선된 헤시안-프리 최적화 기법은 기존의 헤시안-프리 최적화 기법과 비교하여 파형역산의 정확성, 안정성을 제공할 뿐만 아니라 훨씬 빠른 계산 속도를 제공한다.

주요어: 탄성파 파형역산, 뉴턴법, 헤시안-프리 최적화 기법, 음향파 파동방정식

학번: 2012-24168